2005 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

Prepared for the 2005 Methodological Resource Book

Contract No. 283-2004-00022
RTI Project No. 0209009.177.007
Deliverable No. 39

Authors:

Jeremy Aldworth
Kim Ault
Kortnee Barnett-Walker
Lisa Carpenter
Elizabeth Copello
Peter Frechtel
Amy Licata
Bing Liu
Peilan Martin

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

February 2007

2005 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

Prepared for the 2005 Methodological Resource Book

Contract No. 283-2004-00022
RTI Project No. 0209009.177.007
Deliverable No. 39

Authors:

Jeremy Aldworth
Kim Ault
Kortnee Barnett-Walker
Lisa Carpenter
Elizabeth Copello
Peter Frechtel
Amy Licata
Bing Liu
Peilan Martin

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

February 2007

Acknowledgments

This report would not be possible without the guidance and input of staff from the Office of Applied Studies (OAS). In particular, useful comments were provided by Dicy Painter, Joe Gfroerer, and Art Hughes. Special thanks are also due to several current and former RTI International (a trade name of Research Triangle Institute) staff members. Avinash Singh and Ralph Folsom, along with Eric Grau, codeveloped the predictive mean neighborhood (PMN) methodology. Dr. Singh and colleagues provided the text for Appendix B. Larry Myers was instrumental in the implementation of the failure time models for the finer category income imputations and provided most of the content for Sections 9.3.3 and 9.3.4. Jill Ruppenkamp provided valuable input in the writing of the immigrant status variables in Section 5.2. Finally, Richard Hair copyedited the report, Joyce Clay-Brooks formatted the report in preparation for publication, and Brenda Porter helped prepare some of the tables in the appendices.

Table of Contents

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Page

Appendix

List of Tables

1. Introduction

The 2005 National Survey on Drug Use and Health (NSDUH)[1] was implemented using a 50-State multistage cluster design. This design has been in use since the 1999 survey, when this survey was called the National Household Survey on Drug Abuse (NHSDA). Other major changes in the 1999 survey from surveys in previous years included the introduction of computer-assisted interviewing (CAI) methods for both screening households and interviewing selected respondents. The 50-State design has been used since the 1999 survey to allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide direct estimates for eight large States and estimates based on small area estimation (SAE) methods for the remaining States and the District of Columbia.

For the 1999 survey, the introduction of CAI technology was designed to produce more internally consistent data while still allowing the respondent to answer privately by using audio computer-assisted self-interviewing (ACASI) for the more sensitive parts of the interview, such as the drug use modules. Consequently, this ACASI approach allowed the respondent to enter answers to these sensitive questions directly into the computer away from the view of the field interviewer (FI) or any other household members. In addition, the questions were displayed on the screen for the respondent to read, and a recorded voice reading of the questions was provided to the respondent via earphones. Several alternatives to the CAI were evaluated in a field test in 1997, and a smaller pretest of a near-final CAI screening and individual questionnaires was conducted in the summer of 1998 (for details, see Office of Applied Studies [OAS], 2001; Penne, Lessler, Bieler, & Caspar, 1998).

Although the design of the NSDUH survey has not changed significantly since the introduction of CAI in 1999, important methodological changes were introduced in the 2002 survey that affected the estimates from the survey years that followed. In addition to the name change introduced in the 2002 survey, each NSDUH respondent in each survey year since 2002 has received an incentive payment of $30. Also, information from the 2000 U.S. Decennial Census has been used in the NSDUH weighting procedures since the 2002 survey year. Hence, the 2002 survey year is considered the "baseline year," where all trends are measured since that survey year.

This report focuses on the imputation procedures implemented for the 2005 survey. Most of the editing procedures that were applied to the drug, nicotine dependence, income, and health insurance variables, as well as some of the demographic variables requiring imputation (marital status, education, employment status, and immigrant status), are summarized in the 2005 NSDUH editing and coding report (Kroutil, Handley, Suresh, Felts, & Bradshaw, 2007). However, the editing procedures for other demographic variables (age, interview date, birth date, gender, race, and Hispanicity), as well as all of the household composition and proxy variables, are discussed in this document. The criteria used for creating household-level and person-level files, along with eligibility and completeness rules, are discussed in Chapter 2, followed by a

---

[1] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

summary of the implemented imputation procedures in Chapter 3. Chapters 4 and 5 describe the imputation procedures applied to the core and noncore demographic variables, respectively. Chapter 4 also describes editing procedures for age, interview date, birth date, gender, race, and Hispanicity. The drug imputation procedures are discussed in Chapter 6. The imputation procedures for nicotine dependence differed from those used for other variables and are described in Chapter 7. Chapter 8 describes the edits applied to the household roster, the creation and imputation of missing values in the roster-derived household composition variables, and the creation of respondent-level variables with individual roster information. Chapter 9 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 10. Imputations also were conducted in the processing of pair relationships and their accompanying multiplicities for responding pairs, as well as household counts for all households. The procedures used for these imputations were similar to those discussed in this document. However, these imputations are described in a separate document that focuses on the development of household and pair weights (Westlake et al., 2007).

This document also contains 10 appendices, including 3 summaries of the various imputation methodologies used in the current sample. The hot deck is described in Appendix A; the general model used to adjust weights for item nonresponse is discussed in Appendix B; and the methodology developed specifically for NSDUH, the predictive mean neighborhood (PMN) procedure, is described in Appendix C. Respondents had the opportunity to write in responses to some of the drug and demographic questions if they felt the given responses did not apply to them. These responses, called "alpha-specify" or "other-specify" responses, were coded so that the data could be summarized in a meaningful way. A discussion of how this was done for race and Hispanicity is described in Appendix D. (Coding of alpha-specify responses for other variables is summarized by Kroutil et al., 2007.) Models used to assign a single race to multiple race respondents are described in Appendix E. The covariates in each of the imputation models are listed in Appendix F. A summary of the number of respondents who met likeness constraints (i.e., flexible constraints that governed the similarity between donors and recipients) is provided in Appendix G. Appendix H provides details of the vector of predicted means used in the multivariate PMN procedure for employment status, drugs, binary sources of income, and health insurance for various patterns of missing values, in addition to the required logical constraints (i.e., fixed constraints to prevent logical inconsistencies). The quality control measures used in the imputation procedures are summarized in Appendix I. Reasons that interviewers gave for overriding consistency checks in the household roster are presented in Appendix J, along with evaluations of their legitimacy and the resulting actions in the editing of the roster. For the 2005 NSDUH questionnaire specifications for programming, refer to RTI (2006).

2. Household-Level and Person-Level Files

2.1 Sample Design

The population of eligible respondents for the 2005 National Survey on Drug Use and Health (NSDUH)[2] was all civilian, noninstitutionalized residents of the United States (including the District of Columbia) aged 12 or older. As in other recent NSDUHs, this population included residents of noninstitutional group quarters (e.g., homeless shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. Persons excluded from the 2005 survey included those 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 years of age, and active military personnel.

The 2005 survey is the first NSDUH in a coordinated 5-year sample design. Although there is no planned overlap with the 1999-2004 samples, a coordinated design facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. For further details, refer to the 2005 NSDUH sample design report (Morton, Chromy, Hunter, & Martin, 2006).

For the survey, a person was randomly selected for an interview through a four-stage sample selection process. The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. A sample of segments was randomly selected from State sampling (SS) regions during the second stage of sampling.[3] 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. The person-level file was later subset into a smaller data file that contained only respondents who were considered "complete" cases—this file was used for analysis. Refer to Section 2.3 for the definition of complete case. The household-level and person-level files also were utilized in the final creation of the person-level and pair-level analysis weights.

---

[2] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[3] 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 2005 NSDUH sample design report (Morton et al., 2006) contains more information regarding the sample design.

2.2 Dwelling Unit-Level Eligibility and Completeness Criteria

Before proceeding with the fourth stage of sample selection, a set of rules was used to determine whether a DU was eligible to be selected. Examples of ineligible DUs included units defined as "vacant" or "not a primary residency." Eligibility of the DU was recorded in the binary variable DUELIG, where a value of 1 indicated eligibility.

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 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. Because the field interviewers could not obtain an accurate count of DUs from those areas, and because the DUs from the denied-access segments were considered eligible, DU information from the U.S. Bureau of the Census in these areas were 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. This information was updated in the household-level and person-level files.

2.3 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 2.1. 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 to have been selected, but at the time of the interview, the individual could have been determined to be 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 are shown in Table 2.1. The number of eligible persons also is summarized in Table 2.1.

Table 2.1 NSDUH Household and Person Eligibility and Response Rates: 2005

	Selected Dwelling Unit	Eligible Dwelling Units	Completed Screenings	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases
CAI[1]	175,958	146,912	134,055	283,054	83,805	68,373	68,308

[1] 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 gate question and at least 9 of the other 14 gate questions. 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 due to the fact that the computer-assisted interviewing (CAI) instrument routed respondents out of a section if the gate question was not answered. Completeness of eligible individuals was recorded in the binary variable PRCOMP, which had a value of 1 if the respondent was a complete case, and 0 if not. For a summary of the number of completed cases in the 2005 survey, see Table 2.1.

2.4 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 64 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 twenty-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, 25 to 49, and 50 years old 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 also 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 2005 NSDUH sample design report (Morton et al., 2006), 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.[4] This variable had five levels: segment in Core-Based Statistical Areas (CBSA) 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;

[4] Claritas, Inc., is a market research firm headquartered in San Diego, California.

segment in urban area but not in CBSA; and segment in rural area (not in CBSA and not in urban area).

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. Other sampling variables such as DIVISION, SSREGION, GQTYPE, ID, PLACNAME, RURORURB, STATE, STNAME, STUSAB, and QUARTER also were included in the household-level and person-level files.

3. Overview of Item Imputation Procedures

3.1 Introduction

As with most large-scale sample surveys, the 2005 National Survey on Drug Use and Health (NSDUH)[5] faced the problem of analyzing datasets that contained missing responses for some items. In association with this, there were other issues such as inconsistent or invalid responses and violation of skip patterns. Although the instrument was designed to enforce skip patterns, which has reduced inconsistencies relative to paper-and-pencil interviewing (PAPI), and to perform some consistency checks, inconsistent and invalid responses still occurred. These response errors were an obvious source of bias that was considered in the analysis of NSDUH data (Cox & Cohen, 1985).

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, a prediction may not be possible for every record having missing or faulty data (e.g., "cousin" does not clarify the gender of a respondent). The remaining faulty and missing data often are replaced with statistically imputed data.

Since the 1999 survey, NSDUH has been conducted using computer-assisted interviewing (CAI) methods, and the CAI instrument has been the only version used since the 2000 survey. To maintain consistency with surveys since 1999, most of the procedures in the 2005 sample were identical to those used in the previous survey years since 1999 (excluding the 1999 PAPI sample). Each year, however, minor modifications were made to the instrument, which subsequently required adjustments to the imputation procedures, and the 2005 survey was no exception. As in the 2004 survey, the procedure developed specifically for the 1999 survey—the predictive mean neighborhood (PMN) procedure—was applied to most of the variables requiring imputation in the 2005 survey. The only imputations that did not incorporate the PMN method were those used for the nicotine dependence variables, which also were handled differently in the 2004 survey. Table 3.1 provides a brief summary of the types of imputation procedures used for each of the variables imputed in the samples in the 1999 to 2005 surveys.

The vast majority of imputation-revised variables were identified by their names, which were given the prefix "IR." (The imputation-revised employment status variables EMPSTAT4 and EMPSTATY were exceptions to this rule. Although no missing data were possible for gender, the "IR" prefix for IRSEX was maintained for continuity with past years.) Associated

[5] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

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

Variable	1999[1]	2000	2001	2002/2003	2004/2005
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]	MPMN[5]	MPMN	MPMN	MPMN
Hispanic or Latino-Origin Indicator	USHD	UPMN[6]	UPMN	UPMN	UPMN
Marital Status	USHD	MPMN	MPMN	MPMN	MPMN
Hispanic or Latino-Origin Group	USHD	MPMN	MPMN	MPMN	MPMN
Education	USHD	USHD	MPMN	MPMN	MPMN
Employment Status	USHD	USHD	MPMN	MPMN	MPMN
Immigrant Variables	Not imputed	Not imputed	Not imputed	WSHD[7]	UPMN
Health Insurance	MPMN	MPMN	MPMN	MPMN[8]	MPMN
Drug Lifetime Usage	UPMN	MPMN	MPMN	MPMN	MPMN
Recency and Frequency of Use[9]	MPMN	MPMN	MPMN	MPMN	MPMN
Age at First Use	UPMN	UPMN	UPMN	UPMN	UPMN
Age at First Daily Cigarette Use	UPMN	UPMN	UPMN	UPMN	UPMN
Personal and Family Income Binary Variables	MPMN	MPMN	MPMN	MPMN	MPMN
Personal and Family Income Finer Categories	UPMN	UPMN	UPMN	UPMN	UPMN
Nicotine Dependence	Not imputed	Not imputed	Regression	Regression	Regression
Household Size (Roster-Derived Variable)	UPMN	UPMN	UPMN	UPMN	UPMN
Other Household Composition (Roster-Derived) Variables	UPMN	UPMN	UPMN	UPMN	UPMN
Pair Relationship Variables and Multiplicity/Household Counts	PMN[10]	PMN	PMN	PMN	PMN

[1] The 1999 survey year also included a paper-and-pencil interviewing (PAPI) sample. The procedures listed here are from the computer-assisted interviewing (CAI) sample.
[2] "Random" refers to a random assignment within a quarter for the interview date and a random assignment using age and interview date for the birth date.
[3] "None" means that no missing values were encountered after editing, and thus no imputation was necessary. For gender (from the 2002 survey onward) and age, missing values were precluded by design (see Chapter 4).
[4] "USHD" refers to the unweighted sequential hot-deck method of item imputation described in this report (see Appendix A).
[5] "MPMN" refers to the procedure based on the multivariate predictive mean neighborhood model described in this report (see Appendix C).
[6] "UPMN" refers to the procedure based on the univariate predictive mean neighborhood model described in this report (see Appendix C).
[7] "WSHD" refers to the weighted sequential hot-deck method of item imputation described in this report (see Appendix A).
[8] Although MPMN 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 onwards.
[9] "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.
[10] "PMN" refers to the procedure based on the predictive mean neighborhood model that could be univariate or multivariate, depending upon the response variable of the model.

indicator variables, which were identified by the prefix "II," were created to tell the user which values were imputed and which ones were not. For some imputation-revised variables, additional imputation indicators were created with the prefix "II2." These indicators gave more details about the source of the imputed or logically assigned value.

This chapter provides a brief description of PMN, the imputation procedure most used in the 2005 survey, followed by a description of the other procedures used in the survey and a summary of the changes in imputation procedures that occurred between the 2004 and 2005 surveys.

3.2 Overview of PMN Imputation Procedure for the NSDUH Sample

PMN was developed specifically for the 1999 survey. A combination of model-assisted imputation and a random nearest neighbor hot-deck imputation, PMN was implemented for nearly all variables requiring imputation in the 2005 survey (exceptions are shown in Table 3.1).

In general, when large nonresponse occurs, limited donor sets can be used for imputation. For the 2005 survey, to adjust for this sparseness of data, predictive mean modeling was used for the imputation of many of the variables (Table 3.1). The models incorporated sampling design weights[6] with a response propensity adjustment computed to make the item respondent weights representative of the entire sample. The item response propensity model is a special case of the generalized exponential model (GEM),[7] which was developed for weighting procedures. The macro for this model was used to apply the item response propensity model and is described in greater detail in Appendix B. Predicted values (predicted means) were obtained from the models for both item respondents and item nonrespondents. The means of a particular outcome variable were modeled as a function of the predictors (covariates), where these means gave a summary of the effects of covariates on the outcome variable. Unlike the sequential hot-deck imputation method, where values for the covariates were matched through a sorting procedure, the model-based approach used the predicted mean to convert the covariates' effects into a single number. The predicted means, along with other constraints, were used to define the neighborhoods from which donors were randomly selected for the final assignment of imputed values. This assignment was done with either a single predicted mean or several predicted means at once. The method associated with the single predicted mean is called the univariate predictive mean neighborhood (UPMN) method. The multivariate predictive mean neighborhood (MPMN) method is the name associated with the assignment using several predicted means.[8] More details regarding these UPMN and MPMN imputation procedures are provided in Appendix C. For the

---

[6] In the 2005 survey, the final analysis weights were not available in time for imputation processing of almost all variables. The person-level sample design weights were therefore adjusted, using a simple ratio adjustment, to account for nonresponse at the household level. The final analysis weights were used only in the processing of the nicotine dependence variables.

[7] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name for Research Triangle Institute) for weighting procedures.

[8] Although it was often the case that one predicted mean corresponded to one response variable and a vector of predicted means corresponded to several response variables, it was also common practice to (1) assign several values from a single predicted mean (univariate matching, multivariate assignment) or (2) assign a single response value from a vector of predicted means (multivariate matching, univariate assignment). The latter occurred when the response variable was categorical with three or more levels, resulting in a vector of predicted multinomial probabilities, even though only one cell would have a response assigned to it.

types of regression models used for each variable that underwent the PMN imputation procedure, see Table 3.2.

Table 3.2 Regression Models Used for Each Variable Imputed with PMN

Variable	Domain[1]	Type of Regression Model	SAS/SUDAAN Procedure[2,3]
Demographics			
Marital Status	15 years and older	Multinomial Logistic	MULTILOG
Race	All	Multinomial Logistic	MULTILOG
Hispanic Indicator	All	Binomial Logistic	RLOGIST
Hispanic Group	Hispanics	Multinomial Logistic	MULTILOG
Education Level	All	Multinomial Logistic	MULTILOG
Employment Status	15 years and older	Multinomial Logistic	MULTILOG
Immigrant Status: Born-in-U.S. indicator	All	Binomial Logistic	RLOGIST
Immigrant Status: Age of Entry	Not born in U.S.	Simple Linear	REGRESS
Drugs			
Lifetime Drug Use	All	Binomial Logistic	RLOGIST
Recency of Drug Use, "hierarchical" drugs	All lifetime users for past year vs. not past year; all past year users for past month vs. not past month	Binomial Logistic	RLOGIST
Recency of Drug Use, pipes	All lifetime users	Binomial Logistic	RLOGIST
Recency of Drug Use, all other drugs	All lifetime users	Multinomial Logistic	MULTILOG
12-Month Frequency of Drug Use	All past year users	Simple Linear	REGRESS
Daily Drug Use Over Past 30 Days, cigarettes, chewing tobacco, and snuff	All past month users	Binomial Logistic	RLOGIST
30-Day Frequency of Drug Use, cigarettes, chewing tobacco, and snuff	All past month users except those who used daily over the past 30 days	Simple Linear	REGRESS
30-Day Frequency of Drug Use, all other drugs	All past month users	Simple Linear	REGRESS
Age at First Drug Use	All lifetime users	Simple Linear	REGRESS

Table 3.2 Regression Models Used for Each Variable Imputed with PMN (continued)

Variable	Domain[1]	Type of Regression Model	SAS/SUDAAN Procedure[2,3]
Household Composition			
Total Number of Rostered People	All	Poisson	LOGLINK
Total Number of Children Younger Than Age 18	All	Poisson	LOGLINK
Total Number of People Aged 65 or Older	All	Poisson	LOGLINK
Indicator of Whether the Respondent Has Family Members in Household	All	Binomial Logistic	RLOGIST
Income			
Source of Income	All	Binomial Logistic	RLOGIST
Months on Welfare	All respondents who received welfare payments or welfare services in the past year	Simple Linear	REGRESS
Total Income (Binary)	All	Binomial Logistic	RLOGIST
Finer Income Categories	All	Time-to-Event (Survival)	LIFEREG
Health Insurance			
Health Insurance (Old Method)	All	Binomial Logistic	RLOGIST
Health Insurance (Constituent Variables Method)	All	Binomial Logistic	RLOGIST

1 The set of respondents who were included in the model and for whom predicted means were calculated.
2 SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.
3 See RTI (2004) for more information on all procedures except PROC LIFEREG. See SAS Institute (1999) for more information on PROC LIFEREG. PROC LIFEREG is the only SAS procedure in this table. All other procedures are SAS-callable SUDAAN procedures.

Wherever necessary and feasible, additional restrictions were placed on the membership in the hot-deck neighborhoods. These constraints were implemented to make imputed values consistent with preexisting, nonmissing values of the item nonrespondent and to make candidate donors as much like the recipients (the item nonrespondents) as possible. The former are called "logical constraints" and could not be loosened. The latter, called "likeness constraints," could have been loosened if insufficient donors were available to meet the restriction. If more than one likeness constraint was placed on a neighborhood, the restrictions were loosened in a priority order deemed appropriate for the response variable in question.

In the 2005 survey, the variables related to drug use, household composition, income, and health insurance were highly correlated with age. This, along with the desire to expedite the implementation of procedures, made it necessary to separate the model building and final assignments of imputed values for these variables into three distinct age groups. 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 done within the following four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. The age group restriction on the neighborhoods could have been 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 were performed within age groups. This was done to maintain consistency with how the other variables were imputed, and it facilitated easier processing. The same three age groups that were used for drugs were also used for demographics. However, sometimes 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 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 or Latino group, marital status, and immigrant age of entry models.

For the drug variables, there was originally some interest in requiring the donor to be from the same State as the recipient. However, this could not be implemented due to insufficient pools of donors. A different approach was adopted, which was also applied in the 2005 survey: 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 State 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 into one category, respondents from medium usage States into another, and the remainder into a third category. This categorical "State rank" variable was used as one set of covariates in the imputation models. In addition, as another likeness constraint, eligible donors for each item nonrespondent were restricted to be from States with the same level of usage (the same State rank) as the item nonrespondent. A State rank variable was used in a similar manner in the income imputations, both in the modeling and in the hot-deck steps. The three State rank categories were defined in terms of the income level of the States: high-income States, middle-income States, and low-income States. No State rank variables were created for any other variables.

3.3 Other Imputation Procedures Used in the 2005 Survey

Each respondent had a valid age (AGE) and interview date (INTDATE). No imputation was required for these variables. However, sometimes the availability of several alternative values required rules, as outlined in Chapter 4, for selecting the most appropriate values. Missing values for birth date (BRTHDATE) were imputed using a random imputation within the bounds determined by AGE and INTDATE.

The exact date of first drug use was imputed using a random assignment within an interval of possible dates of first use. Each day in the interval was equally likely to be selected. The interval could have been up to a year in length. The date was imputed for almost all lifetime users of each drug, since no respondents were asked for an exact date of first use (though many were asked for the year and month of first use). Chapter 6 provides more details on the algorithm.

The imputation-revised versions of the nicotine dependence variables differed from other imputation-revised variables in three ways: (1) as stated previously in this chapter, PMN was not used to impute missing values; (2) imputed values did not resemble preexisting nonmissing values; and (3) not all missing values were imputed. Weighted least squares regressions were used to obtain continuous predicted means, which were used directly as imputed values. Whereas the nonimputed values were limited to integer values between 1 and 5, imputed values fell anywhere on the continuous scale. Imputations were performed only if the respondent answered at least 16 of the 17 nicotine dependence questions. If the respondent was eligible to answer the nicotine dependence questions, but answered 15 or fewer of them, no attempt was made to replace missing values by imputed values. For these respondents, in the imputation-revised versions of the variables, missing values were still represented as missing values.

3.4 Changes in Procedures from the 2004 Survey to the 2005 Survey

Overall, the changes implemented between the 2004 and 2005 surveys were minor, both in number and in type. No changes to the CAI instrument required corresponding changes in imputation procedures. All changes were procedure enhancements or corrections of minor problems involving both editing and imputation.

In the 2005 survey, new levels were added to the edited recency variables for cigarettes, smokeless tobacco, chewing tobacco, snuff, and cigars, which caused the imputation of these variables to be restricted in new ways. These new levels are described in Table 6.5. In past years, these restrictions were handled by imputation procedures. Conditional predictive means were constructed for these new levels so that in the hot-deck step of PMN, donors were matched to recipients in a more precise manner.

For the first time in the 2005 survey, the "State rank" variable used in imputation processing of lifetime drug use indicators utilized the weighted proportion of lifetime users as the ranking variable, instead of the unweighted proportion of lifetime users. The "State rank" variables used for drug recency of use and income already used weighted values in the ranking in 2004 processing.

The algorithm used in the assignment of the exact date of first drug use was streamlined and documented more thoroughly in the 2005 survey. Also, in Chapter 6 of this document, the method used for the assignment of date of first drug use for parent/child drug pairs is explained more fully.

For the first time in the 2005 survey, the base variable used in the imputation of the lifetime indicator for daily use of cigarettes was an edited variable (CIGDLYMO), instead of a raw variable (CG15). This change in the base variable was done to facilitate the handling of a single respondent whose value for CIGDLYMO differed from the value for CG15, due to edits implemented in an earlier stage of processing.

In the hot-deck program used in the imputation of the lifetime drug use indicators, a new logical constraint was added to cover respondents who were known to have used pain relievers, but whose OxyContin and "other" pain reliever indicators were missing. This situation can occur when respondents respond affirmatively to the question on lifetime use of any of a set of pain relievers shown on a pill identification card (PR04), but fail to select any drugs from the card in the follow-up question (PR04a). In the 2004 survey, the lifetime indicator for "other" pain relievers was created incorrectly. Specifically, all respondents who were lifetime users of any pain reliever were accidentally classified as being lifetime users of other pain relievers. This problem was fixed for the 2005 survey.

New methodology was implemented in the 2005 survey for the generation of seeds for random number generation. In earlier years, either the seed was generated randomly and not recorded, or the same seed was used for every survey year. In the 2005 survey, the seed was generated randomly but recorded, and any reruns used the same seed as in the original run. The advantage of this approach is that any necessary reruns can be done without causing new randomly assigned values to be created; this makes it easier to distinguish changes due to correction or enhancement from changes due to differences in the random number stream.

During the 2005 imputations, a minor error was discovered in the process by which a single donor was selected from a neighborhood of potential donors. In earlier surveys, the "closest" and "furthest" donors in the neighborhood had a lower probability of being selected than the other donors. (Closeness is measured by Mahalanobis distance.[9]) Specifically, given a neighborhood of size n, the closest and furthest donors were each selected with probability $1/(2n - 2)$, and each other donor was selected with probability $1/(n - 1)$. This error was corrected for the 2005 survey.

In the 2004 survey, all imputation procedures used the preliminary analysis weight instead of the final analysis weight. In the 2005 survey, the final analysis weight was used in the processing of the nicotine dependence variables, since it was available at the time the setup programs were run.

A minor error was discovered in Exhibit 7.1 of the 2004 Imputation Report, and in the same exhibit throughout all of the Imputation Reports dating back to the 2001 Imputation Report. Specifically, those respondents who required imputation for cigarette recency, but were known not to be past month users of cigarettes, were incorrectly classified as having their ineligibility imputed. The error was fixed for the 2005 survey, and corrected versions of this table will be included in this report in Chapter 7 for the 2001-2004 surveys. The correction was made in time for inclusion in revised versions of the 2004 Imputation Report, including the HTML version.

[9] See Appendix C for a definition of Mahalanobis distance. A definition also can be found in Manly (1986).

4. Core Demographics

4.1 Introduction

Several demographic characteristics were needed for all respondents in the 2005 National Survey on Drug Use and Health (NSDUH).[10] Core demographic data were collected on both the screener[11] and the questionnaire. Missing values in screener and questionnaire demographic variables were imputed separately for the set of all eligible rostered individuals and for the set of completed respondents (i.e., screener data and questionnaire data were edited and imputed independently).[12] As an initial step, prior to any processing of the data, completed cases were identified. Only these completed cases were included in the subsequent editing, imputation, and analysis of questionnaire data.

The core demographics in the 2005 survey discussed in this report are age, birth date, gender, race, Hispanicity, marital status, and education level (highest grade completed). The only noncore demographic variables imputed were the immigrant variables and employment status. Although the interview date was not classified as a core demographic variable, its editing procedures also are included in this chapter.

Prior to imputation, logical editing was performed on all of these variables. Through the editing process, some missing values were replaced with coded information from the "other-specify" questionnaire responses, thus reducing the amount of statistical imputation required. Noncore information was not used to edit core variables.

After editing, the variables were handled using one of three procedures. For interview date, age, and gender, no statistical imputation was required, because no values were missing after editing. For birth date, 45 respondents had missing values, which were imputed using a random assignment from all possible birth dates that were consistent with the interview date and the age. The missing values in the marital status, race, Hispanicity, and education level variables were imputed using the predictive mean neighborhood (PMN) method. This procedure is described in greater detail in Appendix C. Missing values for the noncore demographic variables, which are discussed in the next chapter, also were imputed using the PMN method.

This chapter describes the editing and imputation procedures used to create the final core demographic variables and interview date for all respondents who were considered "complete cases."[13] A summary of item nonresponse is included for each variable described here.

---

[10] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[11] The "screener" refers to the information about household members obtained at the second stage of sampling in NSDUH, the selection of dwelling units within segments (groups of U.S. Bureau of the Census blocks). The screener information was obtained independently of the questionnaire information.

[12] See the weighting report for the 2005 survey (Chen et al., 2007) for a description of the imputation procedures used for screener demographics for the set of all eligible rostered individuals.

[13] See Chapter 2 for a definition of a "complete case."

4.2 Editing of Demographic Variables

The editing procedures for some of the core demographic variables (marital status and education level) are described in detail by Kroutil, Handley, Suresh, Felts, and Bradshaw (2007) and are briefly summarized here. However, the editing procedures for other core demographic variables (age, birth date, gender, race, and Hispanicity) and interview date are discussed only in this document. Therefore, these variables are described in greater detail in this chapter. For interview date, age, and gender, no imputation was required and the edited variable was considered the final variable to have been used for analysis. There were missing values for birth date, but these values were imputed using a random number, a process that also is described in this section. The variable for birth date that is described in this section also was considered "final." However, the edited variables for marital status, race, Hispanicity, and education level were intermediate variables, since a final imputation as described in Section 4.3 was used to allocate values when data were missing. When a respondent was known to belong to one of several races based on a write-in answer[14] indicating a country of origin, randomly generated numbers were used to allocate the respondent to a particular race. In these cases, the "edited variable" described in this section included these imputed values.

Because information available from the screener could change from survey year to survey year, edits were implemented using only questionnaire data. Screener data were used only in extraordinary circumstances with race imputation models, which are described in Section 4.3.

4.2.1 Interview Date (INTDATE)

Within each module of the questionnaire, after a given module was complete, the time was automatically saved by the computer-assisted interviewing (CAI) instrument. The time for each module was called a "time stamp," and the date portion of the time stamp was called a "date stamp." This information was used to help determine the value for the interview date.

The specific date stamps used to determine the edited interview date (INTDATE) were indicated in the variable EIIDATE. For the labels that define the levels in EIIDATE, if the label indicated that the interview date was set to a particular date stamp, that date stamp was consistent with all subsequent date stamps, unless otherwise indicated. If the interview was set to the end-of-interview date stamp, then that date stamp was consistent with all preceding date stamps except those indicated.

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.)

---

[14] In the section of the questionnaire where the respondent (through the interviewer) selects a race, a respondent can reject the options given and direct the interviewer to provide an alternative answer, also known as a "write-in answer." See Section 4.2.6 for details.

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

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

For a summary of the editing of interview dates, see Table 4.1. As stated above, this information was recorded in the editing indicator variable EIIDATE.

Table 4.1 Interview Date Editing Summary

Value of EIIDATE	Assignment of Interview Date	Frequency	Percent
1	Begin date stamp (all date stamps exist)	68,288	99.96
1.01	Begin date stamp (all date stamps exist except last one)	9	0.01
1.02	Begin date stamp (all date stamps exist up through sedatives)	9	0.01
1.05	Begin date stamp (all date stamps exist up through pain relievers)	1	0.00
1.06	Begin date stamp (all date stamps exist up through inhalants)	4	0.01
3	Tutorial date stamp (begin date stamp is outside quarter)	1	0.00
8	End date stamp (tutorial date stamp is the first occurrence of new date stamp; birthday is between the begin and end date stamp)	1	0.00
8.01	End date stamp (cigarettes date stamp is the first occurrence of new date stamp; birthday is between the begin and end date stamp)	1	0.00
8.12	End data stamp (pain reliever date stamp is the first occurrence of new date stamp; birthday is between the begin and end date stamp)	1	0.00
8.16	End date stamp (noncore demographics date stamp is the first occurrence of new date stamp; birthday is between the begin and end date stamp)	2	0.00
8.17	End date stamp (end date stamp is the first occurrence of new date stamp; birthday is between the begin and end date stamp)	1	0.00

4.2.2 Age

4.2.2.1 Final Edited Age (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 have been different from the age at the end of the questionnaire (NEWAGE). The final age variable, AGE, was determined using these two variables, in addition to three other sources: the age calculated from the final edited interview date (INTDATE) and the raw 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[15] 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

---

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

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 the age variable (AGE) was set to this intermediate age value. There were exceptions, however, as detailed in the following paragraph.

By the time that the interviewer would have reached the roster part of the questionnaire, he or she had multiple opportunities to change the respondent's age stored in the computer in response to consistency checks involving age. 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, the interviewer could have either changed the respondent's age entered in the roster, or overridden the consistency check and provided an explanation as to why the roster age did not match CURNTAGE. If the consistency check for age was overridden, the value for age corresponding to the self may not have matched the intermediate age value described above. However, if the explanations given for overriding the consistency check for age were sufficiently compelling, other evidence pointed to the veracity of the roster age, and the difference between CURNTAGE and the roster age for self was at least 2 years, 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 preinterview screener age matched the roster age.
3. If another member of the household completed an interview, the other household member's roster supported the roster age value.

For a summary of the editing to create AGE for the 2005 survey, see Table 4.2. This information was recorded in the editing indicator variable EIAGE.

Table 4.2 Age Editing Summary

Value of EIAGE	Assignment of Age	Frequency	Percent
1	NEWAGE (consistent with CALCAGE and INTDATE—AGE1)	68,310	99.99
3	NEWAGE (consistent with screener age)	1	0.00
4	CALCAGE (consistent with screener age)	1	0.00
6	NEWAGE (consistent with roster age)	2	0.00
7	CALCAGE (consistent with roster age)	1	0.00
10	Roster age; disagrees with NEWAGE and CALCAGE by at least 2 years, but consistent with screener age, and interviewer specifically indicates that roster age was correct and NEWAGE and CALCAGE were incorrect	3	0.00

4.2.2.2 Recoded Age Categorical Variables (CATAGE, CATAG2, CATAG3)

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.

4.2.3 Birth Date (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. Thus, although a number of cases had missing birth dates, each complete case respondent possessed a current age. When 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), the reported birth month and day were preserved, but 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.

See Table 4.3 for a summary of the birth date editing. This information was recorded in the editing indicator variable EIBDATE.

Table 4.3 Birth Date Editing Summary

Value of EIBDATE	Assignment of Birth Date	Frequency	Percent
1	Reported birth date	68,266	99.92
2	Reported birthday, year from AGE and INTDATE	7	0.01
3	Randomly assigned using AGE and INTDATE	45	0.07

4.2.4 Gender (IRSEX)

As in the 2002-2004 surveys, it was mandatory in the 2005 survey that an interviewer entered 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 2005 survey. However, it was not necessary to create an imputation indicator, since IRSEX and QD01 were exactly equivalent.

4.2.5 Marital Status (MARITAL, EDMARIT)

In the 2005 questionnaire, a single core question (QD07) asked about the respondent's marital status, among respondents 15 or older:

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 et al. (2007). The base variable for creating an imputation-revised version of marital status was called EDMARIT. This was equivalent to MARITAL, except that all legitimate skips were collapsed into a single legitimate skip code (99), and missing values were set to the SAS®[16] missing code (.) so that they could be properly handled by the modeling programs.

4.2.6 Race, Hispanic/Latino Indicator, Hispanic/Latino Group

4.2.6.1 Introduction

In the 2005 questionnaire, two core questions focused on the respondent's race (QD05 and QD05ASIA) and two focused on the respondent's ethnicity[17] (QD03 and QD04). 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 are presented below.

---

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

[17] 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.

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

1 YES
2 NO

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

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

QD05. Which of these groups describes you?

1 WHITE
2 BLACK / AFRICAN AMERICAN
3 AMERICAN INDIAN OR 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 groups describes you?

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

As stated in the guidelines from the Office of Management and Budget (OMB),[18] "Hispanic/Latino" was considered an ethnicity, not a race. However, when given the opportunity to enter a race when the given choices did not apply, many respondents entered "Hispanic" or some Hispanic/Latino group, resulting in a considerable amount of missing data for the race

[18] 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.

question. The final drug-use tables were cross-classified with a variable that combined race and ethnicity. Nevertheless, separate variables were initially created for race and ethnicity, and the race/ethnicity variables used in the tables were derived from these separate variables.

As a result of the confusion between Hispanicity and race, 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 will be discussed in a single section (Section 4.2.6.2).

The procedures used to edit the race and Hispanicity variables in the surveys between 2003 and 2005 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 procedural changes were triggered by the elimination of the QD06 question, which appeared in the survey from 1999-2002. QD06 asked respondents who selected more than one race 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 himself or herself. Details are given in the following sections.

4.2.6.2 Categories Used in Race and Hispanicity Variables

4.2.6.2.1 Race Categories

For editing purposes, the five specific categories in QD05 (white, black/African American, American Indian/Alaska Native, Native Hawaiian, and Other Pacific Islander) and the six specific categories in QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) were combined to produce eleven race categories. Two other categories also were created: "Other Asian" and "Asian nonspecific." Respondents could have chosen almost any subset of these categories. The only impossible subsets were those which 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.

The processing of race accounted for two types of variables: one that included levels for more than one race, and the other, which did not. In addition to the 13 edited single-race categories given above, respondents also could have identified themselves as belonging to a combination of race categories. For some of the variables that accounted for multiple-race responses, these responses were recorded in three levels: more than one race, more than one Asian race, and Native Hawaiian–Other Pacific Islander. Other variables were created that recorded the specific combination of races that was entered. For the variables that did not account for multiple-race responses, multiple-race respondents were allocated to one of the races they selected. This was easily done in survey years prior to 2003, since the response to QD06 (when nonmissing) provided this value. However, with the absence of QD06 since the 2003 survey, a single race was selected from the multiple races chosen in some other manner. The method used for doing this is discussed in Appendix E. A discussion of why this type of variable was needed is given in Section 4.2.6.4.3.

4.2.6.2.2 Hispanic/Latino Categories

With the two additional 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),[19] and they could have chosen more than one category. As with QD05, interviewers could have manually entered the alternative to the choices given, which would have been coded either to some subset of the existing seven categories or set to missing. The other-specify responses to QD05 and/or QD05ASIA, if nonmissing, were consulted if no Hispanic/Latino origin group information was available from QD04. The final imputation-revised Hispanic/Latino group variable, IRHOGRP4, included all seven Hispanic/Latino group levels and a legitimate skip code (99) for respondents who were not Hispanic/Latino.

4.2.6.3 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 Sections 4.2.6.2 and 4.2.6.3, or to some other code that was informative in the final imputation described in Section 4.3. A summary of categories of other-specify codes and how they were handled is given in the following sections. Appendix D gives the individual other-specify codes and more details about how they were handled.

4.2.6.3.1 Mapping of Race Other-Specify Codes

In general, race codes were of four types: (1) directly mapped codes; (2) indirectly mapped codes (these required a quick imputation using a randomly generated number); (3) codes informative for formal imputation procedures; and (4) noninformative codes. The edits following either direct or indirect mapped codes resulted in values that were considered "final." The two other types of codes resulted in incomplete values requiring imputation, and were either informative or noninformative for the formal imputation procedures as described in Section 4.3. Each of the four types of codes is discussed below.

1. Directly Mapped Codes

The directly mapped codes were mapped to one or more of the categories given in the questionnaire (see Section 4.2.6.2). 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, and 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

[19] 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.

as "Han" mapped directly to a category in QD05ASIA ("Chinese") and a response "mestizo" mapped directly to two categories in QD05, "white" and "Native American." Geographic category codes corresponded to a country, where census data indicated a racially homogeneous society. For example, an entry of "Polish" in QD05 mapped to white, since 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 answer 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."

2. 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 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. For example, an entry of "Bolivian" would have a 55 percent chance of being allocated to the American Indian/Alaska Native category, since the latest Bolivian census indicated 55 percent of Bolivians were American Indian/Alaska Native. For countries where the census indicated a small proportion of some indistinct category such as "other" and the randomly generated number indicated an allocation to this proportion, the final race was left to imputation (appropriately constrained based upon the indistinct response). If two or three heterogeneous countries were entered in the other-specify response (e.g., "Bolivian and Peruvian"), the final race was allocated using the following procedure: (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/South Americans (no country listed), and Spanish, which were given codes described under the next subheading, with a final value determined using the formal imputation procedures described in Section 4.3.

3. Codes Informative for Formal Imputation Procedures

Some other-specify responses did not lead to definitive information about the respondent's race. However, the responses were used to limit the final imputation described in Section 4.3. 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.

4. Noninformative Codes

Finally, 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.

4.2.6.3.2 Subsequent Editing of Race Other-Specify Codes

Subsequent to the initial mapping of the 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; (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. Each of these is discussed below.

1. The Final Mapping Depends upon the Source Question

In some cases, 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 have been mapped to "American Indian/Alaska Native" if the other-specify response was in QD05, but "Asian Indian" if the other-specify response was in QD05ASIA. Indirectly mapped codes also could have depended 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 of QD05ASIA. These were carefully examined, but the "Asian" part of the response was always preserved.

2. Responses Given to Both Other-Specify and Non-Other-Specify Categories

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 race categories. For example, if a respondent selected "black/African American" in QD05 and wrote in "black and American Indian," the respondent would be assigned both race 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.[20] 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 peoples in the Polish population.

3. Different Other-Specify Responses Present in at Least Two of QD04, QD05, and QD05ASIA

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.

---

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

4.2.6.3.3 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. If this code appeared in QD04, QD05, or QD05ASIA, the respondent was considered non-Hispanic/Latino regardless of the response to QD03. All other Hispanic/Latino codes either mapped directly to one or more of the seven Hispanic/Latino group categories, or provided no new information (e.g., "Hispanic").

4.2.6.4 Edited Variables, Race

4.2.6.4.1 Individual Race Categories (EDQD051-EDQD0513)

Edited variables were created that correspond to the 13 race categories described in Section 4.2.6.2.1. These variables were called EDQD05*xx*, where *xx* represented a number between 1 and 13, corresponding to each of the 13 categories.

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 EDQD05xx variables had values of 1 or 2, and the level xx was indicated by a directly mapped code in QD04; else

4, if (a) no EDQD05xx 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 had values of 2, 3, or 4 depending on the other-specify codes. [21]

4.2.6.4.2 Broad Categories of Race (EDRACE)

EDRACE summarizes which of four broad race categories (white, black/African American, American Indian/Alaska Native, Asian/Pacific Islander) were identified in QD04, QD05, and QD05ASIA, and it also had levels to indicate how the imputation should have been

---

[21] 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 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.

restricted based on the race of the donor. The first three broad race categories corresponded to EDQD051, EDQD052, and EDQD053, respectively. "Asian/Pacific Islander" was considered to have been identified if any of EDQD054-EDQD0513 was nonmissing. EDRACE was created using the following rules, under five possible scenarios:

Scenario 1: If only one broad race 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/Pacific Islander only), if any of EDQD054 through EDQD0513 were nonmissing.

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

5 (white and black/African American only), if both EDQD051 and EDQD052 were nonmissing; else

6 (white and American Indian/Alaska Native only), if both EDQD051 and EDQD053 were nonmissing; else

7 (white and Asian/Pacific Islander only), if EDQD051 was nonmissing and at least one of EDQD054-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/Pacific Islander only), if EDQD052 was nonmissing and at least one of EDQD054-EDQD0513 were nonmissing; else

10 (American Indian/Alaska Native and Asian/Pacific Islander only), if EDQD053 was nonmissing and at least one of EDQD054-EDQD0513 were nonmissing.

Scenario 3: If three broad race 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-EDQD053 were nonmissing; else

12 (white, black/African American, and Asian/Pacific Islander only), if both EDQD051 and EDQD052 were nonmissing and at least one of EDQD054-EDQD0513 were nonmissing; else

13 (white, American Indian/Alaska Native, and Asian/Pacific Islander only), if both EDQD051 and EDQD053 were nonmissing and at least one of EDQD054-EDQD0513 were nonmissing; else

14 (black/African American, American Indian/Alaska Native, and Asian/Pacific Islander only), if both EDQD052 and EDQD053 were nonmissing and at least one of EDQD054-EDQD0513 were nonmissing.

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

Scenario 5: If none of the broad race 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

missing.

4.2.6.4.3 Broad Categories of Race, No Multiple Race (EDRACEFORMODEL)

Because of the paucity and heterogeneity of multiple-race respondents, imputation models for race did not include a category for more than one race. Instead, predicted means were determined in multinomial logistic models with the following four categories:

1 American Indian/Alaska Native
2 Asian/Pacific Islander
3 Black/African American
4 White

In previous survey years, multiple-race respondents were assigned a single race based on the response to QD06, the multiple-race respondent's "main race." Multiple-race respondents who did not answer QD06 were allocated a "main race" based on an arbitrary priority rule (black/African American, Asian/Pacific Islander, American Indian/Alaska Native, white). Imputation donors were chosen with predicted means for these four categories close to those of the recipient with missing race. A respondent was imputed as being more than one race if the selected donor also identified more than one race.

As in past survey years, an edited variable that did not include a category for more than one race was necessary in the 2005 survey because (1) it was needed to build the imputation models; and (2) it was necessary as a base variable for the final imputation-revised variable that did not include a category for more than one race. However, with the absence of QD06 since the 2003 survey, the respondent did not have an opportunity to indicate a "main race," so a main race had to be assigned probabilistically using models. This edited variable (EDRACEFORMODEL) included the four broad categories given above. Using data pooled across the survey years 2000–2002, a single race was imputed for multiple-race respondents using a series of logistic models. The modeling process is described in detail in Appendix E. Eleven predictive mean models were fit, one for each multiple race category (EDRACE between the values of 5 and 15 inclusive). The parameter estimates from the models were used to impute a "main" or "best" race by the following procedure:

Step 1: Estimate the probability that each respondent would have mentioned each of the broad race categories indicated as their "main" race, using the coefficients from the appropriate predictive mean model.

Step 2: Randomly select one of the broad race categories based on these probabilities.

For example, consider a respondent in the 2005 survey with EDRACE = 5 (white and black/African American only). The covariates included in the model, as described in Appendix E, for respondents with EDRACE = 5 were age, region, race of householder, percentage of owner-occupied households, percentage Asian population, percentage American Indian/Alaska Native population, and percentage black/African-American population. Using the values for these covariates for the 2005 respondent and the parameter estimates from the model, the probability that the respondent would have selected white as his main race could have been estimated. If this probability was estimated at 50 percent, a random imputation was done such that the respondent was assigned white as his main race with probability 50 percent and black/African American as his main race with probability 50 percent.

The assignment of values for EDRACEFORMODEL is summarized below:

EDRACEFORMODEL =

EDRACE, if $1 \leq$ EDRACE ≤ 4; else

randomly imputed main race, if $5 \leq$ EDRACE ≤ 15; else

missing.

4.2.6.4.4 Finer Categories of Race (EDNWRACE)

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

Scenario 1: If only one of EDQD051-EDQD0513 was nonmissing,

EDNWRACE =

16 (Asian nonspecific only), if EDQD0513 was the nonmissing variable; else

xx (one known race category only), where EDQD05xx was the nonmissing variable out of EDQD051-EDQD0512.

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

EDNWRACE =

13 (Native Hawaiian and Other Pacific Islander only), if both EDQD054 and EDQD055 were nonmissing, and all other EDQD05xx variables were missing; else

14 (Asian multiple category), if all of EDQD051-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-EDQD0513 were missing,

EDNWRACE =

15 (More than one race), if EDRACE = 16; else

missing.

4.2.6.5 Edited Variables, Hispanicity

4.2.6.5.1 Hispanic/Latino Indicator (EDHOIND)

The base variable for creating an imputation-revised Hispanic/Latino indicator was EDHOIND, which was created using responses to QD03 and, in rare cases, the other-specify responses to QD04, QD05, and/or QD05ASIA.

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.

Both 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 D.

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

The edited variables EDQD041-EDQD047 were created to match the seven Hispanic/Latino group categories described in Section 4.2.6.2.2: 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 EDQD04xx variables had values of 1 or 2, and the other-specify response from QD05 or QD05ASIA mapped directly to level xx; else

missing.

4.2.6.5.3 Edited Hispanic/Latino Group (EDHOGRP)

The edited variable EDHOGRP was the base variable for creating an imputation-revised Hispanic/Latino group variable. It had seven levels to match the seven Hispanic/Latino group categories described in Section 4.2.6.2.2, plus several other more general Hispanic/Latino levels that could have been 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-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-EDQD047 was nonmissing,

EDHOGRP =

1 (Mexican), if EDQD041 was nonmissing; else

4 (Cuban), if EDQD044 was nonmissing; else

2 (Puerto Rican), if EDQD042 was nonmissing; else

3 (Central or South American), if EDQD043 was nonmissing; else

5 (Dominican), if EDQD045 was nonmissing; else

6 (Spanish), if EDQD046 was nonmissing.

For the multiple-Hispanic/Latino group respondents, an arbitrary 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 the two more Hispanic/Latino group categories since the 2004 survey, resulting in the following order: Mexican, Cuban, Puerto Rican, Central/South American, Dominican, Spanish, and other Hispanic/Latino.

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

EDHOGRP =

EDRACE + 7 (imputation restricted by race), if 1 ≤ EDRACE ≤ 14; else

missing.

4.2.7 Highest Grade Completed (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 done against 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 2005 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 et al. (2007). The base variable for creating an imputation-revised version of marital status 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.

4.3 Demographics Requiring Imputation

Missing values for the demographic variables of completed cases were imputed separately from those of all eligible (screener) rostered individuals. Moreover, almost no screener information was used in the imputation of questionnaire demographics for the completed cases. The exception involved an important covariate in the race imputation model, which is explained in Section 4.3.2. The descriptions that follow discuss the creation of imputation-revised demographic variables. Detailed descriptions of the screener-derived and segment-level[22] covariates used in the imputation models are given in Appendix F.

4.3.1 Marital Status

4.3.1.1 Imputation-Revised Marital Status (IRMARIT)

The variable of interest for marital status was a four-level nominal variable. The four substantive levels of the imputation-revised marital status variable, IRMARIT, were the same as the four answer categories in QD07 (married, widowed, divorced or separated, or never married) and its edited counterparts, MARITAL and EDMARIT, which are described in Section 4.2.5. Respondents younger than 15 were automatically assigned an IRMARIT value of 99, a "legitimate skip" code. The PMN method as applied to the marital status variable is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on multivariate predictive mean neighborhoods (MPMNs).

4.3.1.1.1 Setup for Model Building

Imputations at the hot-deck stage were conducted separately within each of three age groups: 12 to 17, 18 to 25, and 26 or older, though only a single model was fit across all age groups. All respondents with AGE younger than 15 were assigned IRMARIT = 99. Only interview respondents with AGE of 15 or greater were considered as donors.

---

[22] Segments were the second-stage sample units in the multistage 2005 NSDUH sample. Each segment consisted of a set of U.S. Bureau of the Census blocks. Segment-level covariates were defined across the segment in which the respondent's household was located.

An interview respondent was considered an item nonrespondent for marital status if his or her value for EDMARIT was missing. The weights of the item nonrespondents 15 or older were reallocated to the item respondents 15 or older, using an item response propensity model. The weights of the item nonrespondents were redistributed among the item respondents using an item response propensity model. The item response propensity model is a special case of the generalized exponential model (GEM),[23] which is described in greater detail in Appendix B. The covariates in the item response propensity model were census region, gender, population density, age categories, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, percentage of owner-occupied households, and the interaction of age categories and gender.

4.3.1.1.2 Computation of Predicted Means

Using the adjusted weights, the probability of selecting each marital status category (married, widowed, divorced or separated, and never married) was modeled for all age groups together using polytomous logistic regression.[24] The predictors included in the predictive mean model were census region, gender, population density, centered age, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, percentage of owner-occupied households, and the interaction of centered age and gender. These variables were included in both the response propensity and the predictive mean models unless a convergence problem occurred. If this happened, the model was reduced. A summary of the final set of covariates used in the model can be found in Appendix F.

4.3.1.1.3 Assignment of Imputed Values

Separate assignments were performed within each of the three age groups: 15 to 17, 18 to 25, and 26 or older. Respondents aged 12 to 14 were assigned a legitimate skip code 99. The constraints used to select donors are described in the next section.

4.3.1.1.4 Constraints on MPMNs

No logical constraints were used in defining neighborhoods for the marital status variable—only likeness constraints were utilized. In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first constraint required each of the donor's predicted means, as described in Section 4.3.1.1.2, to have been within 5 percent of each of the recipient's three predicted means. The second constraint required donors and recipients to have an age difference of 3 years or less. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the predicted means was removed. See Appendix G for the numbers of respondents meeting each set of likeness constraints on sets of eligible donors.

---

[23] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

[24] All age groups were modeled together because the distributions of the answers for the youngest two age groups were unbalanced, which made it difficult to find convergent models.

4.3.1.1.5 Imputation and Editing Summary for Marital Status

See Table 4.4 for a summary of item nonresponse for marital status (recorded in the variable IIMARIT).

Table 4.4 Marital Status Editing and Imputation Summary

Value of IIMARIT	Assignment of Marital Status	Frequency	Percent
1	From questionnaire	56,969	83.40
3	Statistically imputed	12	0.02
9	Legitimate skip (≤14 years old)	11,327	16.58

4.3.2 Race, Hispanic/Latino Origin Indicator, Hispanic/Latino Group

4.3.2.1 Introduction

As clearly indicated in Section 4.2.6, race and Hispanicity were closely related in the 2005 survey. Moreover, race was used in the imputation of Hispanic/Latino origin, and Hispanicity was used in the imputation of race. As in Section 4.2, which describes editing, the imputation of missing values in the race and Hispanicity variables will be discussed together in this section.

4.3.2.2 Imputation-Revised Race Variables

Sections 4.2.6.4.1 through 4.2.6.4.4 outline the edited variables describing race. Nearly all of these edited variables had imputation-revised counterparts, as shown in Table 4.5. (Some of the individual race category variables were collapsed at the imputation stage.)

All of these variables could have been imputed simultaneously, though the imputations of IRDETAILEDRACE, IRRACE2, and IRNWRACE occurred first, and the imputations of the individual race category variables (IRRACEWH, IRRACEBK, IRRACENA, IRRACENH, IRRACEPI, and IRRACEAS) were subsequently imputed. This was accomplished by assigning values for the individual race category variables using the same donors as in the earlier imputation of IRDETAILEDRACE, IRRACE2, and IRNWRACE. The material will be presented as if the imputation was simultaneous.

Table 4.5 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
EDRACEFORMODEL	IRRACE2
EDNWRACE	IRNWRACE

Whereas their edited counterparts had different codes depending upon the source of the information, the IRRACExx variables were simply binary indicator variables, which were set to 1 if the respondent indicated the given race, and 0 otherwise. The extra information that was contained in the EDQD05*xx* variables was stored in the concomitant IIRACE*xx* variables. The variable IRDETAILEDRACE, which was the only one of these variables not released to the public use and analytic files, gives the same information as the IRRACE*xx* variables, all within a single variable. The final race variable IRRACE2 was a four-level nominal variable: American Indian/Alaska Native, Asian or Pacific Islander, black/African American, and white.[25] This variable has the same levels as IRRACE from previous surveys. The two variables differed in the way they were edited and in the handling of multiple race respondents. Because of the differences, the variable's name was changed. IRNWRACE was a 15-level nominal variable whose levels were the same as the first 15 levels of EDNWRACE.

The imputation-revised race variables were created using an MPMN method for imputation of missing values. The MPMN method as applied to the race variables is explained in detail in the next four subsections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on MPMNs. It should be noted from the outset that the models used in PMN did not have a separate category for multiple-race respondents, due to the small number of these respondents, as well as to their disparate nature. Instead, a model with four broad categories was used: the same broad categories that were found in IRRACE2. Multiple-race respondents in the model were assigned a single race based on the models discussed in Appendix E. They were included in the model-building process as belonging to one of the four broad race categories. Respondents requiring imputation were considered to have been of more than one race if their donor in the hot-deck step of PMN was a multiple-race respondent.

4.3.2.2.1 Setup for Model Building

As with all other variables imputed using PMN methods, the race imputations were conducted separately within age groups. For race and other demographic variables, there were three age groups: 12 to 17, 18 to 25, and 26 or older. The separate age groups were used for ease of processing and consistency with other variables and not because of any strong correlation between age and race. Because all interview respondents were asked the race questions, no subsetting of the data was necessary.

Before predictive mean modeling was implemented, weights were adjusted for item nonresponse to the race questions. (In the 2005 survey, because the final weight adjustments were not completed at the time of the demographic imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.[26]) An interview respondent was considered an item nonrespondent for race if either EDRACEFORMODEL was missing, EDNWRACE was missing or 16, or both. (If the respondent had missing data for either EDRACEFORMODEL or EDNWRACE, he or she also had missing data for the other edited variables in Table 4.5 [EDQD051-EDQD0513 and

---

[25] To collapse the race categories into these four levels, the following categories from QD05 were included in the category "Asian or Pacific Islander": Native Hawaiian, Other Pacific Islander, Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, and other Asian.

[26] In subsequent text, the use of the word "weights" refers to these ratio-adjusted design weights.

EDRACE].) The weights of the item nonrespondents were redistributed among the item respondents using an item response propensity model, one for each of the three age groups. The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in these models included census region, household type (from the screener), age categories (for the respondents aged 26 or older), percentage Hispanic/Latino population, percentage of owner-occupied households, percentage black/African-American population, percentage American Indian/Alaska Native population, and percentage Asian population.

4.3.2.2.2 Computation of Predicted Means

Using the adjusted weights, the probability of selecting each race category was modeled within each age group using polytomous logistic regression.[27] The predictors included in the models were the same as those used in the item response propensity model for race.

The PMN method for race was multivariate, as opposed to univariate, because the predictive mean vector contained more than one element. The three elements in the vector were the predicted probability of being identified with each of the first three race categories (white, black/African American, American Indian/Alaska Native). The probability of being classified as Asian/Pacific Islander was not included, because it was completely defined by the first three elements in the predictive mean vector, being calculated as one minus their sum. A predictive mean vector of three predicted means was created from the polytomous logistic regression model. The covariates in these models included: census region, household type (from the screener), centered age, centered age squared, centered age cubed (oldest age group only), percentage Hispanic/Latino population, percentage of owner-occupied households, percentage black/African-American population, percentage American Indian/Alaska Native population, and percentage Asian population. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

Conditional probabilities were calculated for the few item nonrespondents with EDRACE values of 18 or 19. For details on the computation of these conditional probabilities, see Appendix H.

4.3.2.2.3 Assignment of Imputed Values

For the race questions, the PMN method required the selection of an item respondent who was similar to each item nonrespondent. Specifically, the item respondent "donated" his or her value for the relevant edited variables in Table 4.5 to the item nonrespondent. Most often, the selected item respondent, called the "donor," was randomly chosen from a "neighborhood" of potential donors. The item respondents in this neighborhood were the ones deemed to have been most similar to the given item nonrespondent, who was called the "recipient." Item respondents who were deemed dissimilar to the recipient were discarded from the neighborhood by means of constraints. The predicted means calculated in the previous step were usually considered in these

---

[27] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in the *SUDAAN® Language Manual Release 9.0* (RTI, 2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

constraints. Because multiple variables were considered in the distance measure, "similarity" was defined in terms of the smallest Mahalanobis distance.[28] The PMN methodology is described in more detail in Appendix C. The constraints used for the race variables are described in the next section.

Separate assignments were performed within each of the three age groups. This type of age group-specific assignment was executed for almost all imputation-revised variables in the 2005 survey. If the recipient had missing values for EDRACEFORMODEL and EDNWRACE (as well as the other edited variables in Table 4.5), the donor gave values for all relevant variables to the recipient. In most cases, this ensured consistency between each of the imputation-revised variables. An exception occurred when a respondent listed only one specific category of race, but indicated that he or she was more than one race in the other-specify entry. In these rare cases, the respondent was "more than one race" in IRNWRACE, but only one race was given in the IRRACExx and IRDETAILEDRACE variables.

4.3.2.2.4 Constraints on MPMNs

For the MPMN method, there were two types of constraints: logical constraints and likeness constraints. Logical constraints were not loosened during the search for a donor. Likeness constraints were either loosened or removed if a donor was not found with the given constraints in effect. The logical constraints on the donors for EDRACEFORMODEL and EDNWRACE are listed below:

If the recipient was known to have been Asian (i.e., EDNWRACE = 16), the donor must also have been Asian.

If the recipient had EDRACE = 16 (multiple race, no other information), the donor must have had EDNWRACE = 15.

If the recipient had EDRACE = 17 (nonwhite, no other information), the donor must not have had EDNWRACE = 1.

If the recipient had EDRACE = 18 (white, or both white and American Indian/Alaska Native), the donor must have had EDRACE = 1 or 6.

If the recipient had EDRACE = 19 (not American Indian/Alaska Native, in part or in full), the donor must not have had an EDRACE value of 3, 6, 8, 10, 11, 13, 14, or 15.

In the first attempt to find a neighborhood for each item nonrespondent, a set of likeness constraints was used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that each of the donor's three predicted means (two when the recipients had EDRACE = 19, one when EDRACE = 18), as described in Section 4.3.2.2.2, must have been within 5 percent (within "delta") of each of the recipient's three predicted means. If no potential donors met both of the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed first. If no potential donors met the "delta constraint," the delta constraint also was removed. In addition to these two constraints, a set of likeness constraints concerning the donor's Hispanicity

[28] See Appendix C for a definition of Mahalanobis distance. A definition also can be found in Manly (1986).

were used when the recipient met one of the following conditions. These likeness constraints were never loosened or removed.

> If the recipient was Hispanic/Latino nonspecific (EDHOIND = 1, and all EDQD041-EDQD046 were missing), the donor must have been of Hispanic/Latino origin.
>
> If the recipient selected one or more Hispanic/Latino categories: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spanish (EDHOIND = 1, and one or more EDQD041-EDQD046 were nonmissing), the donor must have had an EDHOGRP value equal to one of the Hispanic/Latino groups mentioned by the recipient. For example, if the recipient chose Mexican and Central or South American, the donor must have had EDHOGRP = 1 or 3.
>
> If the recipient had EDRACE = 20 (non-Hispanic Mexican), the donor must have been Mexican (but the donor could have been Hispanic/Latino or non-Hispanic/Latino).

The likeness constraints for the race variables, along with the number of respondents meeting each set of likeness constraints on sets of eligible donors, are listed in Appendix G.

4.3.2.2.5 Imputation and Editing Summary for Race

To differentiate the final imputed values from nonmissing values, a concomitant indicator variable, IIRACE2, indicated how the levels of IRRACE2 were derived. Table 4.6 gives the levels for the indicators of the individual race category variables (IIRACExx). The levels for IRRACE2 are provided in Table 4.7. The 15-level race variable, IRNWRACE, also had a concomitant indicator variable. Table 4.8 summarizes the levels of IINWRACE, the concomitant indicator variable for IRNWRACE. No indicator variable was created for IRDETAILEDRACE.

Table 4.6 IRRACExx Editing and Imputation Summary

Value of IIRACExx	Assignment of IRRACExx	*xx* = WH (white)		*xx* = BK (black/African American)		*xx* = NA (American Indian/Alaska Native)	
		Freq.	Pct.	Freq.	Pct.	Freq.	Pct.
1	Directly selected/not selected	66,327	97.10	66,556	97.44	66,473	97.31
2	From other-specify	223	0.33	27	0.04	91	0.13
3	From census data	78	0.11	45	0.07	64	0.09
4	Statistically imputed	1,680	2.46	1,680	2.46	1,680	2.46
Value of IIRACExx	**Assignment of IRRACExx**	***xx* = NH (Native Hawaiian)**		***xx* = PI (Pacific Islander)**		***xx* = AS (Asian)**	
		Freq.	**Pct.**	**Freq.**	**Pct.**	**Freq.**	**Pct.**
1	Directly selected/not selected	66,627	97.54	66,626	97.54	66,353	97.14
2	From other-specify	1	0.00	2	0.00	268	0.39
3	From census data	0	0.00	0	0.00	7	0.01
4	Statistically imputed	1,680	2.46	1,680	2.46	1,680	2.46

Table 4.7 IRRACE2 Editing and Imputation Summary

Value of IIRACE2	Assignment of IRRACE2	Frequency	Percent
1	From single QD05 response	64,085	93.82
2	Logically assigned from alpha-specify response	415	0.61
3	Single race imputed from multiple responses	1,998	2.92
4	Single race assigned with census data from country of origin	66	0.10
5	Multiple races assigned with census data, single race imputed	64	0.09
6	Statistically imputed (unrestricted)	25	0.04
7	Statistically imputed (restricted)	1,655	2.42

Table 4.8 IRNWRACE Editing and Imputation Summary

Value of IINWRACE	Assignment of IRNWRACE	Frequency	Percent
1	From QD05 response(s)	65,958	96.56
2	Logically assigned from alpha-specify response(s)	556	0.81
3	Assigned with census data from country of origin	128	0.19
4	Statistical imputation of "Asian" into finer categories	6	0.01
5	Statistically imputed (unrestricted)	25	0.04
6	Statistically imputed (restricted)	1,635	2.39

4.3.2.3 Imputation-Revised Hispanic/Latino Indicator (IRHOIND)

As with the imputation-revised race variables, a PMN method was used for the Hispanic/Latino indicator. However, because there was only one element in the predictive mean vector in this case, a univariate predictive mean neighborhood (UPMN) method was used. The PMN method as applied to the Hispanic/Latino indicator is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on UPMNs.

4.3.2.3.1 Setup for Model Building

As with imputations for other race variables, the imputations for the Hispanic/Latino indicator were conducted separately within the three age groups: 12 to 17, 18 to 25, and 26 or older. The separate age groups were used more for ease of processing and consistency with other variables rather than due to any strong correlation between age and Hispanic/Latino origin. Because all interview respondents were asked the question about Hispanic/Latino origin, no subsetting of the data was necessary.

As for the race variables, weights were adjusted for item nonresponse to the Hispanic/Latino origin question, QD03, using item response propensity models, one for each age group. (Weights were defined in a similar manner to the way weights were determined for other demographic variables. Details on how the weights were defined can be found in Section 4.3.2.2.1.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The potential covariates in the item response propensity model were census region, imputation-revised race, age categories (for the 26 or older age group), percentage Hispanic/Latino population, percentage of owner-occupied households, percentage black/African-American population, percentage American Indian/Alaska Native population, and percentage Asian population.

4.3.2.3.2 Computation of the Predicted Means

Using the adjusted weights, the probability of an affirmative response to the Hispanic/Latino origin question was modeled within each age group using logistic regression. The predictors included in the models were census region, imputation-revised race, household type, centered age, centered age squared, centered age cubed, imputation-revised marital status, percentage Hispanic/Latino population, percentage of owner-occupied households, percentage black/African-American population, percentage American Indian/Alaska Native population, and percentage Asian population. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

4.3.2.3.3 Assignment of Imputed Values

Separate assignments were performed within each of the three age groups: 12 to 17, 18 to 25, and 26 or older. The constraints used to select donors are described in the next section.

4.3.2.3.4 Constraints on UPMNs

No logical constraints were used in defining neighborhoods; only likeness constraints were utilized. In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that the donor's predicted mean, as described in Section 4.3.2.3.2, must have been within 5 percent of the recipient's predicted mean. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed. A donor was found for every item nonrespondent using this method. Therefore, no further loosening of constraints was necessary. See Appendix G for the numbers of respondents that met each set of likeness constraints on sets of eligible donors.

4.3.2.3.5 Imputation and Editing Summary for Hispanic/Latino Origin

Less imputation was required for the Hispanic/Latino indicator than for the race variables. Table 4.9 summarizes item nonresponse for the Hispanic/Latino indicator. This information was recorded in the variable IIHOIND.

Table 4.9 Hispanic/Latino Indicator Editing and Imputation Summary

Value of IIHOIND	Assignment of IRHOIND	Frequency	Percent
1	From questionnaire	68,146	99.76
2	From alpha-specify responses	6	0.01
3	Statistically imputed	156	0.23

4.3.2.4 Race and Hispanicity Recodes Used in Subsequent Processing

The imputation-revised race (IRRACE2) and imputation-revised Hispanic/Latino indicator (IRHOIND) variables were used to create several additional race/ethnicity variables. One of these was used in the subsequent processing of imputation-revised variables. Since the 2003 survey, this variable (RACE2) had four levels: non-Hispanic/Latino white, non-Hispanic/Latino black/African American, Hispanic/Latino, and non-Hispanic/Latino other. However, it was similar to a variable created in previous survey years (RACE) from IRRACE and IRHOIND. RACE had the same levels as RACE2. Other variables were created from IRNWRACE and IRHOIND that were used extensively in the production of tables (NEWRACE1 and NEWRACE2).

4.3.2.5 Imputation-Revised Hispanic/Latino Group (IRHOGRP4)

4.3.2.5.1 Introduction

Due to the fact that two additional Hispanic/Latino group categories (Dominican and Spanish) were added to QD04 in the 2004 questionnaire and these additions also appeared in the 2005 survey, a final imputation-revised Hispanic/Latino group variable IRHOGRP4 was created to differentiate from IRHOGRP3. This variable, IRHOGRP3, was the imputation-revised Hispanic/Latino group variable created prior to the 2004 survey. With the added Hispanic/Latino group category "Dominican," there were very few respondents who classified themselves as non-Dominican Caribbean. Therefore, the level Hispanic/Latino Caribbean that was present in IRHOGRP3 was eliminated from IRHOGRP4, and collapsed into the "Other" Hispanic/Latino group category. The edited variable EDHOGRP, described in Section 4.2.6.5.3, 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. It was created using an MPMN method similar to the method for IRMARIT. The predictive mean vector had only three elements associated with the first three levels of EDHOGRP: the predicted probabilities of the interview respondent being Puerto Rican, Mexican, and Cuban. Using only three predicted means made the computation of both predicted means and Mahalanobis distances more feasible.[29]

The PMN method as applied to the Hispanic/Latino group variable is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on MPMNs.

4.3.2.5.2 Setup for Model Building

All respondents with IRHOIND = 2 were automatically assigned IRHOGRP4 = 99 and were excluded from the item response propensity model, the predictive mean model, and the set of potential donors. In contrast to the other demographic variables, imputations were not conducted separately within age groups. This was done for two reasons. First, with combined age groups, the models were likely to be better because none of the response categories were sparsely populated. Second, only respondents with IRHOIND = 1 were eligible to be donors, so it was

[29] The ordering of the levels of IRHOGRP4 differed from the questionnaire and from EDHOGRP. The levels were rearranged after all the imputation programs were complete.

necessary to keep all age groups in the same dataset to ensure donor pools that were sufficiently large.

An interview respondent was considered an item nonrespondent for Hispanic/Latino group if his or her value for EDHOGRP was missing. The weights of the item nonrespondents were then redistributed among the item respondents using an item response propensity model (see Appendix B for the more general GEM), and covariates included census region, imputation-revised race, gender, age categories, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, percentage of owner-occupied households, and the interaction of age categories and gender.

Starting in the 2004 survey, respondents who indicated multiple Hispanic/Latino groups also were excluded from the model-building process. In past survey years, if a respondent indicated multiple Hispanic/Latino groups, the single Hispanic/Latino group selected that was determined depended upon a priority rule: Mexican, Cuban, Puerto Rican, Central/South American, Caribbean Islander, and other Hispanic/Latino. Even though this priority rule was arbitrary, respondents who were assigned a Hispanic/Latino group based on this priority rule had been used in the Hispanic/Latino group models since the 1999 survey. Because the Hispanic/Latino group model did not include a separate level for multiple Hispanic/Latino groups, respondents with multiple Hispanic/Latino groups were not considered item respondents in both the item response propensity model and the predictive mean model.

4.3.2.5.3 Computation of Predicted Means

Using the adjusted weights, the probability of selecting each of the first three Hispanic/Latino group categories (according to EDHOGRP) was modeled for all age groups together, using polytomous logistic regression. The predictors included in the predictive mean model were the same as the predictors used in the response propensity model except for age-related covariates where the continuous version of age were used in the model. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

4.3.2.5.4 Assignment of Imputed Values

All age groups were aggregated in this step, for the reasons given in Section 4.3.2.5.2. The constraints used to select donors are described in the next section.

4.3.2.5.5 Constraints on MPMNs

No logical constraints were used in defining neighborhoods; only likeness constraints were utilized. In the first attempt to find a neighborhood for each item nonrespondent, three likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that if the recipient had $8 \leq$ EDHOGRP ≤ 21, the donor's IRDETAILEDRACE value had to indicate a subset of the race categories mentioned by the recipient. For example, if the recipient had EDHOGRP = 13 (Hispanic/Latino group missing, and the only races mentioned were white and American Indian/Alaska Native), the donor must have had IRDETAILEDRACE of 1 (white only), 3 (American Indian/Alaska Native only), or 6 (white and American Indian/Alaska Native only).

The third likeness constraint stated that each of the donor's three predicted means, as described in Section 4.3.2.5.1, must have been within 5 percent of each of the recipient's three predicted means. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed. If still no donor was found, the constraint on the predicted means also was removed. The constraint involving race was never loosened or removed. See Appendix G for the numbers of respondents that met each set of likeness constraints on sets of eligible donors.

4.3.2.5.6 Imputation and Editing Summary for Hispanic/Latino Group

To differentiate the final imputed values from nonmissing values, a concomitant indicator variable, II2HOGR4, gave the source of information for IRHOGRP4. The levels of II2HOGR4 are summarized in Table 4.10. A variable that gave somewhat less information, IIHOGRP4, also was created to give the source of information for IRHOGRP4. The levels of IIHOGRP4 and II2HOGR4 were identical to the Hispanic/Latino group indicator variables created prior to the 2004 survey. Table 4.10 shows how the levels of II2HOGR4 mapped to those of IIHOGRP4. As previously stated in Section 4.2.6.5.3, a priority rule[30] was used to determine what group a respondent belonged to if he or she gave more than one response. The variable II2HOGR4 recorded these cases, whereas IIHOGRP4 merely considered these cases a "response from questionnaire."

Table 4.10 Hispanic/Latino Group Editing and Imputation Summary

Value of II2HOGR4	Assignment of IRHOGRP4	Frequency	Percent	Level of IIHOGRP4
1	Single Hispanic/Latino group from questionnaire	9,392	13.75	1
2	Single Hispanic/Latino group from alpha-specify response(s)	106	0.16	2
3	Single Hispanic/Latino group determined from multiple responses	253	0.37	1
4	Statistically imputed (unrestricted), or IRHOIND imputed to 2	156	0.23	3
5	Statistically imputed (restricted by race)	40	0.06	4
9	Legitimate skip (respondent was not Hispanic/Latino – nonimputed)	58,361	85.44	9

4.3.2.6 Imputation-Revised Multiple Hispanic/Latino Group (IRHOGRPM)

4.3.2.6.1 Introduction

As in past survey years, respondents were asked to choose the Hispanic/Latino group(s) that best described them in QD04. They were allowed to select more than one Hispanic/Latino group and also could write in an answer in the QD04 other-specify. For the respondents from more than one Hispanic/Latino group, a priority rule (see Section 4.2.6.5.3) was used to determine the final Hispanic/Latino group for the respondent. In surveys prior to the 2004

[30] Amended slightly from previous surveys, the priority rule for the 2005 survey was the following: Mexican, Cuban, Puerto Rican, Central/South American, Dominicans, Spanish, and other Hispanic/Latino.

NSDUH, there was no single variable that could report this information. Since the 2004 survey, a multiple Hispanic/Latino groups variable (IRHOGRPM) was created to capture the information when a respondent identified multiple Hispanic/Latino groups.

4.3.2.6.2 Imputation and Editing Summary for Multiple Hispanic/Latino Group

The imputed variable IRHOGRPM was created based on the imputation-revised variables IRHOIND, IRHOGRP4, and edited variables EDQD041-EDQD047 as described in Sections 4.3.2.3, 4.3.2.5, and 4.2.6.5.2, respectively. For the Hispanic/Latino group nonrespondents, the values of EDQD041-EDQD047 from the donors in IRHOGRP4 were used in place of the missing Hispanic/Latino group categories for the recipients. The first seven levels of IRHOGRM that represented the single Hispanic/Latino group respondents were Puerto Rican only, Mexican only, Cuban only, Other Hispanic/Latino only, Central or South American only, Dominican only, and Spanish only. Level eight represented the respondents from multiple Hispanic/Latino groups. A legitimate skip code 99 was assigned to the non-Hispanic/Latino group respondents:

IRHOGRPM =

99, if IRHOIND = 2; else

1 to 7, if IRHOIND = 1 and only one of EDQD041-EDQD047 was selected; else

8, if IRHOIND = 1 and more than one of EDQD041-EDQD047 were selected.

The source information for IRHOGRPM was recorded in its indicator variable IIHOGRPM, which is summarized in Table 4.11.

Table 4.11 Multiple Hispanic/Latino Group Editing and Imputation Summary

Value of IIHOGRPM	Assignment of IRHOGRPM	Frequency	Percent
1	Single or multiple Hispanic/Latino groups from questionnaire	9,635	14.11
2	Single or multiple Hispanic/Latino groups from alpha-specify response(s)	108	0.16
3	Multiple Hispanic/Latino groups from questionnaire and alpha-specify responses	8	0.01
4	Statistically imputed (unrestricted), or IRHOIND imputed to 2	156	0.23
5	Statistically imputed (restricted by race)	40	0.06
9	Legitimate skip (respondent was not Hispanic/Latino – nonimputed)	58,361	85.44

4.3.2.7 Hispanic/Latino Group Recodes Used in Subsequent Processing

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

4.3.3 Core Education

4.3.3.1 Imputation-Revised Highest Grade Completed (IREDUC)

As with the marital status, race, and Hispanic/Latino group variables, the predictive mean modeling for the highest grade completed variable was done using polytomous logistic regression. The base edited variable EDEDUC has 17 substantive levels (the same as in QD11), but these were collapsed into fewer levels for ease of modeling. For respondents aged 12 to 17, the predictive mean vector had four elements. For the other two age groups (18 to 25 and 26 or older), the predictive mean vector had three elements. The PMN method as applied to the highest–grade-completed variable is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on MPMNs.

4.3.3.1.1 Setup for Model Building

The imputations for the highest grade completed variable in the hot-deck stage were conducted separately within the three age groups: 12 to 17, 18 to 25, and 26 or older. Because all interview respondents were asked this question, no subsetting of the data was necessary. Two of these age groups were aggregated for the modeling stage: 18 to 25 and 26 or older.

Weights were adjusted for item nonresponse to the highest–grade-completed question, QD11. The covariates in the item response propensity model (see Appendix B for the more general GEM) were census region, imputation-revised race, gender, age categories (except in the 12-to-17 age group), the interaction of age categories and gender (except in the 12-to-17 age group), percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, and percentage of owner-occupied households.

4.3.3.1.2 Computation of Predicted Means

For ease of modeling, the 17 substantive levels of EDEDUC were collapsed into fewer levels. For respondents aged 12 to 17, the response variable in the predictive mean model had five levels: less than elementary school (EDEDUC = 1, 2, 3, 4, or 5), elementary school (EDEDUC = 6 or 7), middle school (EDEDUC = 8 or 9), some high school (EDEDUC = 10 or 11), and high school (EDEDUC = 12 or higher). For respondents aged 18 or older, the response variable had four levels: less than high school (EDEDUC < 12), high school (EDEDUC = 12), some college (EDEDUC = 13, 14, or 15), and college or higher (EDEDUC = 16 or 17).

Using the adjusted weights, the probability of the respondent having each level of the response variable was modeled using polytomous logistic regression. The respondents aged 12 to 17 years old were modeled separately from the two older age groups. For the youngest age group, the predictors included in the model were census region, imputation-revised race, gender, centered age, centered age squared, centered age cubed, the interaction of centered age and gender, the interaction of centered age squared and gender, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska

Native population, percentage Asian population, and percentage of owner-occupied households. For the other two age groups, the predictors included in the model were census region, imputation-revised race, gender, centered age, centered age squared, centered age cubed, the interaction of centered age and gender, the interaction of centered age squared and gender, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, percentage of owner-occupied households, and imputation-revised marital status. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

4.3.3.1.3 Assignment of Imputed Values

Separate assignments were performed within each of the three age groups: 12 to 17, 18 to 25, and 26 or older. The constraints used to select donors are described in the next section.

4.3.3.1.4 Constraints on MPMNs

No logical constraints were used in defining neighborhoods for the education level variable; only likeness constraints were utilized. For the two youngest age groups, three likeness constraints were used in the first attempt to find a neighborhood for each item nonrespondent. The first required the donor to have been the same age as the recipient. The second stated that the donor must have lived in the same segment as the recipient. The third likeness constraint stated that the donor's predicted means, as described in Section 4.3.3.1.2, must have been within 5 percent of the recipient's predicted means. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed. If potential donors still were not found, the delta constraints were removed. For the oldest age group, the constraints were the same except that the constraint on the donor's age was not applied. See Appendix G for the numbers of respondents meeting each set of likeness constraints on sets of eligible donors.

4.3.3.1.5 Imputation and Editing Summary for Highest Grade Completed

Table 4.12 summarizes item nonresponse for the highest grade completed variable. This information was recorded in the variable IIEDUC.

Table 4.12 Highest Grade Completed Editing and Imputation Summary

Value of IIEDUC	Assignment of IREDUC	Frequency	Percent
1	From questionnaire	68,294	99.98
3	Statistically imputed	14	0.02

4.3.3.2 Education Records

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 18 or older, some college and 18 or older, college graduate and 18 or older, or 12 to 17 years old).

5. Noncore Demographics

5.1 Introduction

For the 2005 National Survey on Drug Use and Health (NSDUH),[31] missing values were imputed in two sets of variables in the noncore demographics module: the immigrant status and employment status variables. Additionally, the core demographics that were imputed in the 2005 survey are discussed in Chapter 4.

For immigrant status, two imputation-revised variables (IRBORNUS and IRENTAG2) were created using the edited variables BORNINUS and ENTRYAG2 as base variables.[32] Respectively, these variables recorded whether a respondent was born in the United States, and, if not, the age of entry into the United States. The name of the age-of-entry variable (ENTRYAG2) was changed from previous surveys due to changes in the questionnaire that are described in Section 5.2. ENTRYAG2 is analogous to the variable IRENTAGE, which had been used for previous surveys. The final goal was to create a data file containing variables that would have indicated whether respondents could have been included in incidence analyses based on their immigrant status.

The variables describing current employment status were determined from 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 and whether that week was atypical. The employment status questions were asked of only respondents aged 15 or older. A single imputation-revised variable, EMPSTATY, was created from the series of employment status questions. Unlike other imputation-revised variables, for historical reasons this variable was not preceded by an "IR" prefix. However, it was accompanied by imputation indicators that did have the requisite "II" prefix, II2EMSTY and IIEMPSTY.

Respondents who either worked during the week preceding the interview or said they had a job were asked to write in the industry for which they worked, their occupation, and their main duties at work. Edited versions of the responses to some of these questions are discussed in a separate document (Kroutil, Handley, Suresh, Felts, & Bradshaw, 2007). Even though responses were edited, missing values were not imputed.

---

[31] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[32] 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.

5.2 Immigrant Status

The edited immigrant status variables used to create IRBORNUS and IRENTAG2 are described in Section 5.2.1. The edited variable BORNINUS, the base variable used for creating IRBORNUS, was derived from the questionnaire questions QD14 and QD15, and is described in Section 5.2.1.1. Whereas the indicator of whether the respondent was born in the United States did not change from previous surveys, the determination of the length of time non-U.S.-born respondents had lived in the United States changed since the 2004 survey. In surveys prior to 2004, the length-of-time information was obtained from a single question (QD16) in categorical increments. However, since the 2004 survey, this information was obtained from three questions (QD16a, QD16b, and QD16c), from which a continuous amount of time in the United States was obtained. The edited variables LIVUS1YR, LIVUSYRS, and LIVUSMOS were derived from these questionnaire questions and were consolidated into a single variable LNGTHLIV. The edited age of entry variable (ENTRYAG2) was derived from LNGTHLIV and was used as the base variable for IRENTAG2. The variables LIVUS1YR, LIVUSYRS, and LIVUSMOS are discussed in Section 5.2.1.2; LNGTHLIV and CONTAGE are discussed in Section 5.2.1.3; and ENTRYAG2 is discussed in Section 5.2.1.4. Imputation-revised immigrant status variables had been imputed using the weighted sequential hot-deck (WSHD) method since the 2002 survey. However, partly because of the changes in the questionnaire, inconsistency problems were not easily resolved using the WSHD method when creating IRENTAG2. To alleviate this problem and to promote consistency with how imputations were conducted with other variables in NSDUH, imputations on the immigrant status variables IRBORNUS and IRENTAG2 have been conducted using the predictive mean neighborhood (PMN) method since the 2004 survey, which is discussed in detail in section 5.2.2. The variables IRBORNUS and IRENTAG2 were subsequently used to create recoded variables for the purposes of analysis. The recoded Hispanic/Latino group variable HISPGRP2, which prior to the 2004 survey was used specifically for the imputation of missing values in the immigrant status variables, was no longer required in the PMN method. Nevertheless, to maintain consistency with previous surveys, HISPGRP2 was created since the 2004 survey, and is described in Section 5.2.3.

5.2.1 Edited Immigrant Status Variables

5.2.1.1 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 a separate document (Kroutil et al., 2007).

5.2.1.2 Length of Time Lived in the United States (LIVUS1YR, LIVUSYRS, and LIVUSMOS)

As previously stated, the 2005 survey questions recording the length of time that a non-U.S.-born respondent had lived in the United States changed from surveys prior to 2004. In surveys prior to the 2004 NSDUH, the length of time that non-U.S.-born respondents had lived in the United States was obtained from a single question (QD16). Since the 2004 survey, however, respondents were 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 QD16b 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 et al. (2007).

5.2.1.3 Continuous Age (CONTAGE) and Continuous Length of Time Lived in the United States (LNGTHLIV) for non-U.S.-born Respondents

In order 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. Since QD16b and QD16c were designed to be mutually exclusive, the edited variables LIVUSMOS and LIVUSYRS were combined to create the continuous estimate of how many years a non-U.S.-born respondent had lived in the United States, LNGTHLIV. 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, their 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. Since interview date and birth date, as described in Chapter 4, had no missing values, CONTAGE also had no missing values. A legitimate skip code 999 was assigned to the respondents who were born in the United States for both LNGTHLIV and CONTAGE.

5.2.1.4 Age of Entry (ENTRYAG2)

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.

5.2.2 Imputation-Revised Immigrant Status Variables

5.2.2.1 Imputation-Revised Born-in-U.S. Indicator (IRBORNUS)

As with all other demographic variables requiring imputation, except birth date, the PMN method was used to impute missing values in the born-in-U.S. indicator variable. Since the born-in-U.S. indicator was a single dichotomous discrete variable, the assignment of imputed values was univariate. The univariate PMN (UPMN) procedure is described in Appendix C.

The UPMN procedure, as applied to IRBORNUS, is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on UPMNs.

5.2.2.1.1 Setup for Model Building

Imputation of missing values in the born-in-U.S. indicator was conducted within three age categories: 12 to 17, 18 to 25, and 26 or older. The separate age groups were used more for ease of processing and consistency with other variables rather than due to any strong correlation between whether a respondent was born in the United States and age. Because all interview respondents were asked the question whether they were born in the United States, no subsetting of the data was necessary.

If a valid response ("yes" or "no") was provided for the born-in-U.S. measure, the person was defined as an item respondent. The weights were adjusted for item nonresponse using item response propensity models, one for each age group. The item response propensity model is a special case of the generalized exponential model (GEM),[33] which is described in greater detail in Appendix B. The covariates in these models included gender, age categories, gender by age category interactions, imputation-revised race/ethnicity, imputation-revised education level, imputation-revised employment status, imputation-revised marital status, percentage of owner-occupied households, metropolitan statistical area, and census region.

5.2.2.1.2 Computation of Predicted Means

After the weight adjustment, the probability of an affirmative response to the question whether a respondent was born in the United States was modeled within each age group using logistic regression. The predictors included gender, centered age, centered age squared, centered age cubed, gender by centered age squared interaction, gender by centered age cubed interaction, imputation-revised race/ethnicity, imputation-revised education level, imputation-revised employment status, imputation-revised marital status, percentage of owner-occupied households, metropolitan statistical area, and census region. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

---

[33] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

5.2.2.1.3 Assignment of Imputed Values

Separate assignments were performed within each of the three age groups: 12 to 17, 18 to 25, and 26 or older. The constraints used to select donors are described in the next section.

5.2.2.1.4 Constraints on UPMNs

No logical constraints were used in defining neighborhoods—only likeness constraints were utilized. In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that the donor's predicted mean, as described in Section 5.2.2.1.2, must have been within 5 percent of the recipient's predicted mean. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed. If a potential donor could still not be found, the delta constraints were removed. In the 2005 survey, a donor was found for every item nonrespondent using this method. Therefore, no further loosening of constraints was necessary. See Appendix G for the number of respondents that met each set of likeness constraints on sets of eligible donors.

5.2.2.1.5 Imputation and Editing Summary for Born in the United States

Table 5.1 summarizes item nonresponse for the born-in-U.S. variable. The source information was recorded in the indicator variable IIBORNUS.

Table 5.1 IRBORNUS Editing and Imputation Summary

Value of IIBORNUS	Assignment of IRBORNUS	Frequency	Percent
1	From questionnaire	68,279	99.96
2	Logically assigned	6	0.01
3	Statistically imputed	23	0.03

5.2.2.2 Imputation-Revised Immigrant Age of Entry (IRENTAG2)

The PMN method was utilized for imputing missing values in the variable recording the age of entry into the United States of non-U.S.-born respondents. It followed the same general procedures as the imputation of other demographic variables. A linear regression model was fitted using a logit transformation of the respondent's age of entry as the response variable. Because the immigrant's age of entry was a single continuous variable, a UPMN method, as described in Appendix C, was used in the imputation-revised age of entry assignment. The PMN method as applied to IRENTAG2 is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on UPMNs.

5.2.2.2.1 Setup for Model Building

All respondents who were born in the United States (IRBORNUS = 1) were assigned a legitimate skip code (IRENTAG2 = 999) and were excluded from the item response propensity model, the predictive mean model, and the set of potential donors. Imputations of missing values in the age-of-entry variable were not conducted separately within age groups because the number of non-U.S.-born respondents was too small to support quality models and sufficient donor pools in three separate age groups.

An interview respondent was considered as an item nonrespondent for the age-of-entry variable if the edited variable ENTRYAG2 had missing data. The weights were adjusted for item nonresponse using item response propensity models to match the entire non-U.S.-born population. The item response propensity model is a special case of GEM, which is described in greater detail in Appendix B. The covariates in these models included gender, age categories, gender by age category interactions, imputation-revised race/ethnicity, imputation-revised education level, imputation-revised employment status, imputation-revised marital status, percentage of owner-occupied households, metropolitan statistical area, and census region.

5.2.2.2.2 Computation of Predicted Means

The predicted means for an immigrant's age of entry was estimated using a linear regression model. To control the upper and lower bound of predicted means for age of entry, it was necessary to perform a logit transformation to the response variable. The response variable in the model was the immigrant age at entry as a proportion of the continuous version of current age CONTAGE, as described in Section 5.2.1.3. The expression of the proportion is $P_i = Y_i/N_i$, where $Y_i = Age\ at\ Entry_i$ and $N_i = Continuous\ Age_i$ (CONTAGE). After the weight adjustment, the following empirical logit transformation was used as the response variable in a weighted linear univariate regression:

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

This transformation was nearly equivalent to the standard logit transformation:

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

which was not used because it might be unstable for respondents who entered the country at their current age. Variables included in the regression model were gender, centered age, centered age squared, centered age cubed, gender by centered age squared interaction, gender by centered age cubed interaction, imputation-revised race/ethnicity, imputation-revised education level, imputation-revised employment status, imputation-revised marital status, percentage of owner-occupied households, metropolitan statistical area, and census region. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F.

5.2.2.2.3 Assignment of Imputed Values

The assignment was performed on the full sample and was not separated into age categories, for reasons given in Section 5.2.2.2.1. The constraints used to select donors are described in the next section.

5.2.2.2.4 Constraints on UPMNs

Two logical constraints and two likeness constraints were utilized in the definition of neighborhoods for IRENTAG2. Both logical constraints involved the respondent's age. One required that the donor's age of entry be less than the recipient's current age. The other logical constraint required the difference between the recipient's current age and the donor's age of entry to be less than 1 year if the recipient lived in the United States for less than 1 year (as indicated by QD16a), or the difference had to be greater than 1 if the recipient lived in the United States for more than 1 year.

In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that the donor's predicted mean, as described in Section 5.2.2.2.2, must have been within 5 percent of the recipient's predicted mean. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the segment of the potential donor was removed. If a potential donor could still not be found, the delta constraints were removed. See Appendix G for the number of respondents that met each set of likeness constraints on sets of eligible donors.

5.2.2.2.5 Imputation and Editing Summary for Immigrant Age of Entry

The associated indicator variable for the imputation-revised immigrant age of entry was IIENTAG2. Table 5.2 summarizes item nonresponse for the age-of-entry variable.

Table 5.2 IRENTAG2 Editing and Imputation Summary

Value of IIENTAG2	Assignment of IRENTAG2	Frequency	Percent
1	From questionnaire	7,217	10.57
2	Logically assigned	6	0.01
3	Statistically imputed (including those imputed to IRBORNUS = 2)	43	0.06
9	Legitimate skip (BORNINUS = 2)	61,042	89.36

5.2.3 Recoded Hispanic/Latino Group Variable (HISPGRP2)

Prior to the 2004 survey, when the weighted sequential hot-deck method was used, two variables—HISPGRP2 and AGEADULT—were created specifically to aid in the imputation of missing values in the immigrant status variables. These variables were no longer needed when the PMN method was used starting with the 2004 survey. The variable AGEADULT, which has not been produced since the 2004 survey, would have been equivalent to CATAG3, which is described in Chapter 4. The variable HISPGRP2 was created in the 2005 survey in the same way as in previous surveys, having been derived from the variable IRHOGRP4. Some of the levels in IRHOGRP4 were collapsed to generate a more condensed version of the Hispanic/Latino group variable. As a result, HISGRP2 had four levels: 1 = Puerto Rican (IRHOGRP4 = 1), 2 = Mexican (IRHOGRP4 = 2), 3 = Other Hispanic/Latino (IRHOGRP4 = 3, 4, 5, 6, or 7), and 4 = Non-Hispanic/Latino (IRHOGRP4 = 99).

5.3 Current Employment Status

The edited employment status variables used to create EMPSTATY are described in Section 5.3.1. Section 5.3.1.1 discusses the edited variables JBSTATR and WRKHRSUS. Section 5.3.1.2 discusses the creation of EDEMPY, the base variable for imputation. Sections

5.3.2 and 5.3.3 discuss the imputation procedure for EMPSTATY, and Section 5.3.4 discusses the creation of EMPSTAT4, a recoded version of EMPSTATY.

5.3.1 Edited Employment Status Variables

5.3.1.1 JBSTATR and WRKHRSUS

The main edited variable used to summarize the respondent's current work situation was JBSTATR, which was subsequently used to create EMPSTATY. This edited variable combined information from QD26, QD29, QD30, QD31, QD32, and QD33. The categories for JBSTATR are shown in Table 5.3. 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 et al. (2007).

Table 5.3 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		

5.3.1.2 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 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 = 11-15, 290, or 299; else

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

missing.

5.3.2 Imputation-Revised Employment Status (EMPSTATY)

Missing values in the edited employment status variable EDEMPY were replaced with imputed values using a multivariate predictive mean neighborhood (MPMN) procedure. This procedure is described in greater detail in Appendix C. The MPMN method was applied to employment status variables for the first time in the 2001 survey. It was enhanced in the 2002 survey to account for partial knowledge of employment status. The imputation procedure for employment status in the 2005 survey was similar to the procedures that have been used since the 2002 survey.

The MPMN method as applied to the employment status variable is explained in detail in the next four sections: setup for model building, computation of predicted means, assignment of imputed values, and constraints on MPMNs.

5.3.2.1 Setup for Model Building

Similar to the imputations that were performed on other demographic variables, imputations for employment status variables in the hot-deck stage of the PMN method were conducted separately within the same three age groups: 12 to 17, 18 to 25, and 26 or older. All respondents with AGE < 15 were assigned EMPSTATY = 99. Only interview respondents with AGE ≥15 were used in the models or were considered as donors. At the modeling stage of PMN, two of these age groups were aggregated: 15 to 17 and 18 to 25.

An interview respondent was considered an item nonrespondent for employment status if his or her value for EDEMPY = 5 (employed, part-time versus full-time unclear) or missing. The weights of the item nonrespondents 15 or older were reallocated to the item respondents 15 or older. (In the 2005 survey, the final analysis weights were used if they were available. However, because the final weight adjustments were not completed at the time of the demographic imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.[34]) The item response propensity model is a special case of GEM, which is described in greater detail in Appendix B. Respondents aged 15 to

[34] In subsequent text, the use of the word "weights" will refer to the ratio-adjusted design weights.

25 were modeled separately from respondents aged 26 or older.[35] The initial set of covariates in the two models were the same: census region, imputation-revised race, gender, age categories, the interaction of age categories and gender, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, and percentage of owner-occupied households.

5.3.2.2 Computation of Predicted Means

Using the adjusted weights, the probability of selecting each employment status category (employed full-time, employed part-time, unemployed, and other) was modeled using polytomous logistic regression.[36] The predictors included in the model for the respondents aged 15 to 25 were census region, imputation-revised race, gender, centered age, centered age squared, the interaction of centered age and gender, the interaction of centered age squared and gender, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, and percentage of owner-occupied households. The predictors included in the model for the respondents aged 26 or older were census region, imputation-revised race, gender, centered age, centered age squared, centered age cubed, the interaction of centered age and gender, the interaction of centered age squared and gender, percentage Hispanic/Latino population, percentage black/African-American population, percentage American Indian/Alaska Native population, percentage Asian population, percentage of owner-occupied households, and imputation-revised marital status. The number of covariates was reduced if convergence or stability problems occurred in the model-fitting process. A summary of the final set of covariates used in the model can be found in Appendix F. The predictive mean vector used in the imputation procedure had three elements (three predicted probabilities) corresponding to the first three levels of EDEMPY.

5.3.2.3 Assignment of Imputed Values

The imputations were performed separately within each of three age groups: 15 to 17, 18 to 25, and 26 or older. All constraints used to select donors are described in the next section.

5.3.2.4 Constraints on MPMNs

One logical constraint was used in defining neighborhoods for the employment status variable: if the recipient had EDEMPY = 5, the donor must have been employed either part-time or full-time (EDEMPY = 1 or 2).

---

[35] The 15- to 17-year-old respondents were separated from the 18- to 25-year-old respondents in the stage where final imputed values were assigned. This separating of age groups was done because these two age groups had very different work patterns. However, in both the response propensity models and the predictive mean models, these two age groups were combined. This combining of age groups was done because there were an insufficient number of 15- to 17-year-old working respondents to get viable models

[36] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in the *SUDAAN® Language Manual Release 9.0* (RTI, 2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

Conditional probabilities were used to take advantage of the partial information that was available. Recipients with EDEMPY = 5 were known to be employed. Instead of the usual three predicted means using the model's predicted probabilities directly, a single predicted mean was derived using a conditional probability, which was the probability that the recipient was employed full-time given that the respondent was employed. See Appendix H for more details on missingness patterns for employment status.

In addition to the logical constraint, three likeness constraints were used. In the first attempt to find a neighborhood for each item nonrespondent, the donor's age was required to be within 4 years of the recipient's age; the donor was required to live in the same segment as the recipient; and each of the donor's three predicted means (one predicted mean for recipients with EDEMPY = 5), as described in Section 5.3.2.2, were required to be within 5 percent of each of the recipient's three predicted means. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the donor's segment was removed. If still no donors were found, the delta constraints were removed. See Appendix G for the numbers of respondents meeting each set of likeness constraints on the sets of eligible donors.

5.3.3 Imputation and Editing Summary for Employment Status

See Table 5.4 for a summary of item nonresponse for employment status. The table shows the values of both the detailed imputation indicator II2EMSTY and the simpler indicator IIEMPSTY.

Table 5.4 EMPSTATY Editing and Imputation Summary

Assignment of EMPSTATY	Frequency	Percent	Value of IIEMPSTY	Value of II2EMSTY
From Questionnaire	56,938	83.35	1	1
Statistically Imputed (Unrestricted)	19	0.03	3	3
Statistically Imputed (Restricted to Full-Time or Part-Time)	24	0.04	3	4
Legitimate Skip (Respondent Was 12-14 Years Old)	11,327	16.58	9	9

5.3.4 Imputation-Revised Employment Status Recode (EMPSTAT4) and Indicators (II2EMST4 and IIEMPST4)

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).

The same relationship was held between both II2EMSTY and II2EMST4 and IIEMPSTY and IIEMPST4. II2EMSTY was equivalent to II2EMST4, and IIEMPSTY was equivalent to IIEMPST4 for respondents younger than 15 or older than 17. For respondents aged 15 to 17, II2EMST4 = IIEMPST4 = 9.

6. Drugs

6.1 Introduction

Major changes were introduced in the imputation procedures for the drug use variables in the computer-assisted interviewing (CAI) sample of the 1999 National Household Survey on Drug Abuse (NHSDA), which was renamed the National Survey on Drug Use and Health (NSDUH) in 2002.[37] In particular, for the CAI sample of the 1999 survey, a new imputation methodology (i.e., predictive mean neighborhood [PMN]) was developed specifically for NSDUH. This methodology is a combination of weighted regression and nearest neighbor hot-deck imputation, where the hot deck is random whenever possible.[38] Its application to the drug use variables for the 2005 survey was similar to that of previous survey years, as is explained in the following sections.

This chapter describes how the PMN technique was applied to the drug use variables. In some cases, imputations were required because the respondent did not answer a given question. However, other responses were altered in the editing process due to inconsistencies. In these cases, the original response was set to missing, or, in the case of recency of use, a specific recency was edited to a more general recency that was consistent with other responses, and determination of the specific recency was left to imputation. For example, a recency-of-use response might have been edited to past year usage, where past-month versus past-year-but-not-past-month use could have been determined by imputation. These editing processes are summarized by Kroutil, Handley, Suresh, Felts, and Bradshaw (2007).

The models for these imputations, which are described in detail in the following sections, were either weighted logistic regression models (binomial or multinomial) or weighted multiple linear regression models with the response variable appropriately transformed. Using the PMN technique, the predicted means from these models were used to determine neighborhoods, from which donors were randomly selected for the final assignment of imputed values. (If no donors were available within a very small distance of the recipient's predicted mean, the donor with the closest predicted mean was chosen.) The neighborhoods were created based on a single predicted mean (a univariate predictive mean neighborhood [UPMN]) or using several predicted means at once (a multivariate predictive mean neighborhood [MPMN]). Even if the neighborhood was constructed from a univariate predicted mean, the assignment of imputed values could have been either univariate or multivariate. The members of the neighborhood were restricted to satisfy two types of constraints: "logical constraints" and "likeness constraints." Constraints that made the imputed values consistent with preexisting values of other variables were called logical constraints and were required for the candidate donor to have been a member of the neighborhood. Likeness constraints were implemented to make donors and recipients as much alike as possible. Although logical constraints could not have been loosened, likeness constraints

[37] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[38] The nearest neighbor hot deck is described in detail in Appendix A.

could have been loosened if they forced the donor pool to have been too sparse. Details of these PMN imputation procedures are provided in Appendix C.

In the 2005 survey, because drug use was highly correlated with age, and to facilitate easier 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.[39]

Although statistical imputation of the drug use variables could not have proceeded separately within each State due to insufficient pools of donors, information about the State of residence of each respondent was incorporated in the modeling and hot-deck steps in the sample. States were classified into three drug usage categories within each age group: States with high usage of a given drug were placed in one category, States with medium usage into another, and the remainder into a third category. Respondents were then assigned values for a three-level "State rank" variable, depending on their State of residence. The indicator variables resulting from this categorical State rank variable were used as covariates in the imputation models. In addition, for all of the drug use measures, eligible donors for each item nonrespondent were restricted, if possible, to be from States with the same level of usage (the same State rank) as the item nonrespondent. The definition of "level of usage" (i.e., what measure of usage was used to categorize the States) depended on the drug use measure being imputed.

As with the CAI instruments used in the 1999 through 2004 surveys, the 2005 survey had different drugs and drug use measures than are found in pre-1999 surveys. Table 6.1 summarizes the drugs and drug use measures that were imputed and whether the imputations were univariate or multivariate. If no character is present in the box in Table 6.1, then no information regarding that particular drug use measure was available for the given drug.

6.2 Hierarchy of Drugs and Drug Use Measures

The first step in the imputation process was to determine the order in which drugs and drug use measures were to be modeled so that drugs and drug use measures earlier in the sequence could have been used, if applicable, as covariates for models fitted later in the sequence. Because the gate questions in the 2005 survey were the basis for all subsequent drug data, it was necessary that the imputation of missing values for lifetime drug use for all drugs preceded imputations of all other drug use measures. These lifetime use indicators were temporary in the sense that they were manifested within the drug recency and frequency-of-use variables, but they were not delivered themselves. The hierarchy of models for drugs for the lifetime usage models is discussed in Section 6.3.1.

Once all the lifetime usage indicators had been determined, the imputations of the remaining measures proceeded. Where indicated in Table 6.1, a multivariate imputation was implemented within each drug for recency of use, 12-month frequency of use, 30-day frequency

[39] The modeling procedures were done separately within each of the three age groups regardless of the response variable.

of use, and binge drink 30-day frequency (alcohol only). For a given drug, recency of use[40] was included in the model for frequency of use, 12-month frequency of use was included in the model for 30-day frequency, and 30-day frequency of use of alcohol was included in the model for the binge drink frequency variable.

Table 6.1 Drugs and Drug Use Measures, Univariate Versus Multivariate Imputation[1]

	Drug Use Measure						
Drug	**Lifetime Usage**	**Recency of Use**	**12-Month Frequency of Use**	**30-Day Frequency of Use**	**Binge Drink Frequency**	**Age at First Use**	**Age at First Daily Use**
Cigarettes	✓✓	✕		✕		✓	✓
Smokeless Tobacco[2]	✓✓	✕✕		✕✕		✓✕	
Cigars	✓✓	✕		✕		✓	
Pipes	✓✓	✓					
Alcohol	✓✓	✕	✕	✕	✕	✓	
Inhalants	✓✓	✕	✕	✕		✓	
Marijuana	✓✓	✕	✕	✕		✓	
Hallucinogens[3]	✓✓	✕✕	✕✕	✕✕		✓✕	
Pain Relievers[4]	✓✓	✕✕	✕✕			✓✕	
Tranquilizers	✓✓	✕	✕			✓	
Stimulants[5]	✓✓	✕✕	✕✕			✓✕	
Sedatives	✓✓	✕	✕			✓	
Cocaine and Crack	✓✓	✕✕	✕✕	✕✕		✓✕	
Heroin	✓✓	✕	✕	✕		✓	

✓ Univariate neighborhood; univariate assignment of imputed values.
✓✓ Multivariate neighborhood across all lifetime drug use variables; multivariate assignment of imputed values across all lifetime drug use variables.
✕ Multivariate neighborhood across recency of use, 12-month frequency of use where applicable, 30-day frequency of use where applicable, and the 30-day binge drink frequency variable (alcohol only); multivariate assignment of imputed values across measures.
✕✕ Multivariate neighborhood across recency of use, 12-month frequency of use where applicable, and 30-day frequency of use where applicable; multivariate assignment of imputed values across these measures and across certain drugs (see Section 6.5.5.1.3).
✓✕ Univariate neighborhood and multivariate assignment of imputed values (see Section 6.6.1.7).
[1] If no character is present, then no information regarding that particular drug use measure was available for the given drug.
[2] Includes chewing tobacco and snuff.
[3] Includes LSD, PCP, and Ecstasy.
[4] Includes OxyContin.
[5] Includes methamphetamines.

Finally, age at first use was required to have been consistent (in a number of ways) with the other measures (see Section 6.6). Hence, age at first use was imputed after the imputation for

[40] Missing values were replaced by imputed values in these recency- and frequency-of-use variables. The imputed values were provisional since the final values were not known until the multivariate imputation, after the completion of the modeling.

the other measures was completed.[41] The following sections describe the imputation procedures for each drug use measure.

Some of the rows of Table 6.1 refer to both a general drug category and one or more subcategories. In the remainder of this chapter, to highlight the relationship between them, these drugs 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 "subgate" questions. However, they also could appear in separate modules. The parent-child drug combinations included smokeless tobacco (parent) and chewing tobacco and snuff (children); cocaine (parent) and crack (child); hallucinogens (parent) and LSD, PCP, and Ecstasy (children); pain relievers (parent) and OxyContin (child); and stimulants (parent) and methamphetamines (child). Smokeless tobacco differs from the other parent drugs in that data were not collected on this drug. Respondents were asked about only the two child drugs (chewing tobacco and snuff). Any measures of smokeless tobacco can be considered as recoded variables because they were not directly imputed. Table 6.2 illustrates all the drugs in parent/child relationships and the data that were gathered on them.

Table 6.2 Drugs in a Parent/Child Relationship

Parent Drug	Child Drug(s)	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
Stimulants	**Methamphetamines**	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 6.3.7.3.

6.3 Imputing Lifetime Drug Use Indicators

As with the 1999 through 2004 surveys, the 2005 survey implemented automatic routing through the questionnaire. Using a series of gate questions, the instrument asked the respondent

[41] For cigarettes, both age at first use and age at first daily use had to have been consistent with the other measures. Hence, age at first use was imputed after the other measures, followed by the imputation of age at first daily use.

whether he or she had ever used a number of drugs in 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 him or her to the next module. 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.

6.3.1 Hierarchy of Drugs

Because PMN was used for the lifetime usage imputations, a drug hierarchy was required, the use of which was motivated in general for PMN as described in Appendix C. Experience from past survey years has indicated a substantial correlation between lifetime drug use indicators. Although models were built using respondents with complete data across all the drugs, predicted means were calculated for both item respondents and nonrespondents for lifetime use. When calculating the predicted means for the lifetime usage of a given drug for respondents who did not answer all the lifetime usage questions, a predictor value could have been missing. Hence, it was sometimes necessary to use imputed lifetime usage values. These imputed values were provisional, since the final imputed lifetime usage indicators were not known until the final multivariate imputation, after the completion of the modeling.

Therefore, the first step in the imputation of lifetime indicators was to determine the order in which the drugs would be modeled, where drugs later in the sequence would have more predictors in their models. The order in which the lifetime indicators of use were imputed is shown in Table 6.3.

6.3.2 Setup for Model Building and Hot-Deck Assignment

Once the hierarchy of drugs was established, the next step was to define respondents, nonrespondents, and the item response mechanism. As stated earlier, imputations for all drug use measures were conducted separately within the three age groups: 12 to 17, 18 to 25, and 26 or older. For an individual to have been considered a lifetime-use item respondent, he or she must have had complete data within each age group 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; methamphetamines; stimulants other than methamphetamines; and sedatives. Response propensity adjustments were then computed for each age group to make the item respondent weights representative of the entire sample. (Because the modeling of the final weight adjustments was not completed at the time of the drug imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.)[42] An adjustment was calculated that reallocated weights from item nonrespondents to item respondents. Because item respondents were defined across all drugs, this adjustment was computed only once per age group and then used in the modeling of lifetime use for all drugs. The item response propensity model is a

[42] In subsequent text, the use of the word "weights" will refer to the ratio-adjusted design weights.

special case of the generalized exponential model (GEM),[43] which is described in greater detail in Appendix B.

Table 6.3 Lifetime Indication of Use ("Gate") Questions (in Order of Imputation)[1]

Drug	Question(s)
Cigarettes	CG01
Smokeless Tobacco[2]	CG17, CG25
Cigars	CG34
Pipes	CG42
Alcohol	AL01
Inhalants	IN01a, IN01b, IN01c, IN01d, IN01e, IN01f, IN01g, IN01h, IN01i, IN01j, IN01l
Marijuana	MJ01
Hallucinogens[3]	LS01a, LS01b, LS01c, LS01d, LS01e, LS01f, LS01h
Pain Relievers[4]	PR01, PR02, PR03, PR04, PR04a, PR05
Tranquilizers	TR01, TR02, TR03, TR04, TR04a, TR05
Stimulants[5]	ST01, ST02, ST03, ST04, ST04a, ST05
Sedatives	SV01, SV02, SV03, SV04, SV04a, SV05
Cocaine	CC01
Crack	CK01
Heroin	HE01

[1] Follow-up questions also were considered in the lifetime imputation.
[2] Includes chewing tobacco (CG17) and snuff (CG25).
[3] Includes LSD (LS01a), PCP (LS01b), and Ecstasy (LS01f).
[4] Includes OxyContin (option 12 in PR04a).
[5] Includes methamphetamines (ST01).

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 through 11 different questions if 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 inhalant. 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,

[43] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

stimulants, sedatives, and smokeless tobacco. To have been considered a lifetime nonuser of a drug module with multiple gate questions, the respondent had to have answered "no" to all of the gate questions. If none of the gate questions in a drug module was answered affirmatively, but some of the gate questions were unanswered, the individual was considered a nonrespondent for that module.

6.3.3 Sequential Model Building

Starting with cigarettes, the probability of lifetime use of each drug was modeled for item respondents, within each age group, using the nonresponse-adjusted weights. Logistic regression[44] was used to determine the parameter estimates. Because the interest was in only the estimation of the predicted mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. The predictors in each model included lifetime use of drugs already imputed; centered age[45]; centered age squared; centered age cubed; gender; race/ethnicity; first-order interactions of age, race/ethnicity, and gender; a three-level State rank variable (incorporating the proportion of lifetime users of the drug of interest in the respondent's State of residence); population density; and census region.[46] For age groups 18 or older, the variables for marital status, education level, and employment status also were included. For a complete summary of the lifetime use imputation models, see Appendix F.

6.3.4 Computation of Predicted Means and Creation of Univariate Predictive Mean Neighborhoods

Using the parameter estimates from the probability of lifetime usage model for a given drug, predicted probabilities of use were computed for both item respondents and nonrespondents. These predicted values were then used to temporarily impute a value for each nonrespondent, using the UPMN imputation method described in Appendix C. Although models were built using respondents with complete data across all drugs, predicted probabilities were required for all respondents. In order to use lifetime usage of a given drug as a predictor for a drug later in the sequence, it was therefore necessary to utilize these temporary imputed values in cases where the original lifetime usage indicator was missing. If possible, provisional donors were chosen with predicted means within the delta of the recipient,[47] where the value of delta varied depending on the value of the predicted means, which in this case were predicted probabilities of lifetime use. In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and it was defined as 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5 and allowed a tighter delta for predicted probabilities close to

---

[44] SAS®-callable SUDAAN® was used to fit the binomial and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in RTI (2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

[45] The covariate age was centered within each age group in order to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[46] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

[47] "Delta" refers to the value that defined the neighborhood of donors that were "close" to the item nonrespondent. The difference between the predicted mean of the item nonrespondent and the predicted means of the item respondents in the neighborhood must have been less than delta. See Appendix C for more details.

0 or 1. The range of values for delta across various predicted probabilities is shown in Table 6.4. If no donors were available with predicted means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predicted mean was chosen.

Table 6.4 Values of Delta for Various Predicted Probabilities of Lifetime Use

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

6.3.5 Assignment of Provisional Imputed Values

Subject to the constraints described in the next section, separate assignments of provisional values were performed within each of the three age groups. The final lifetime imputations were multivariate across lifetime drug use variables and are further described in Section 6.3.8.

6.3.6 Constraints on Univariate Predictive Mean Neighborhoods

In a general UPMN imputation, the neighborhood is restricted by two types of constraints: (a) logical constraints (which cannot be loosened) to make imputed values consistent with a nonrespondent's preexisting nonmissing values of other variables, and (b) likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible. The next paragraph discusses the likeness constraints and the order in which they were loosened or removed.

As with all other drug use measures, neighborhoods for lifetime use indicators were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, and 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. A small delta also could have been considered a likeness constraint, which could have been loosened by enlarging or removing delta. As previously stated, if no donors were found in the delta, as defined in Section 6.3.4, the neighborhood was then abandoned and the donor with the predicted mean closest to the recipient was chosen.[48] If possible, donors and recipients were required to be from States with the same level of usage of a given drug (the State rank, as defined in the introduction of this chapter), where the level of usage was defined in terms of the proportion of a given State's residents who were lifetime users of the drug. 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 hallucinogens was missing, but the recipient was known to be a lifetime nonuser of LSD, then the donor must also have been a lifetime nonuser of LSD.) If insufficient donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned and the donor with the closest predicted mean was chosen; and (2)

[48] Although using neighborhoods is important for the calculation of the variance due to imputation, methods to account for donor-predicted means differing greatly from recipient-predicted means had not yet been devised at the time these imputations were implemented.

both the State rank and child drug lifetime 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. Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. Even in those cases, however, the imputation was performed so that no logical constraints were necessary, as discussed in Section 6.3.7.

6.3.7 Multivariate Assignments

Although the methodology for determining the nearest neighbor neighborhood was univariate in terms of the predicted probability of lifetime use, peculiarities associated with drugs in parent/child pairs sometimes required the assignment step to have been multivariate. These drugs are discussed in the following sections.

6.3.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

Many respondents who indicated lifetime use of smokeless tobacco seemed to have been 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 nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted probability of lifetime use. Missing values for chew and/or snuff were replaced with the values from a donor within this neighborhood. For individuals missing the lifetime usage indicator for either chew or snuff but not both, only the missing value was replaced. However, for individuals missing both chew and snuff, both lifetime usage indicators were replaced by values from the same donor. No logical constraints were necessary in the assignment step. This was due to the fact that chew and snuff were assigned values independently, and then combined at the end to form a final lifetime usage indicator for smokeless tobacco.

6.3.7.2 Cocaine and Crack

Because cocaine and crack were in distinct modules in the 2005 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. This neighborhood was defined in terms of the deltas shown in Table 6.4, which were based on the predicted probabilities of lifetime use for both cocaine and crack. An item respondent was eligible to have been a donor for a given item nonrespondent if his or her predicted probability of lifetime cocaine use was within delta of the item nonrespondent's cocaine-predicted probability and his or her predicted probability of lifetime crack use was within delta of the item nonrespondent's crack-predicted probability. This was true regardless of whether the item nonrespondent was missing only crack or both crack and cocaine.[49] Once the neighborhood was defined, missing values for crack and/or cocaine were replaced with the values from a donor within this neighborhood. For individuals missing a lifetime usage indicator for only crack, but not both crack and cocaine, only the missing value was replaced. However, for individuals missing both crack and cocaine, both lifetime usage indicators were replaced by values from the same donor. It is important to note that it would not have been possible for a respondent to have been missing a value for cocaine but not crack because a crack user is, by definition, also a cocaine user. For this reason, no logical constraints were necessary.

6.3.7.3 Hallucinogens (LSD, PCP, Ecstasy, and Other Hallucinogens), Pain Relievers (OxyContin and Other Pain Relievers), and Stimulants (Methamphetamines and Other Stimulants)

The hallucinogens, pain relievers, and stimulants modules all included multiple gate questions (called "subgate questions"), and some of the substances referenced in the subgate questions were child drugs. For hallucinogens, there were three child drugs: LSD, PCP, and Ecstasy. For pain relievers, there was one child drug: OxyContin. For stimulants, there was also one child drug: methamphetamines.

Predicted probabilities were calculated for the parent drugs, and these probabilities were used to determine neighborhoods for each group of drugs. An "other" category was created by combining all the other subgate questions with the exception of the ones referring to the child drugs. In the final assignment step, lifetime usage indicators were assigned for LSD, PCP, Ecstasy, and "other" hallucinogens; OxyContin and "other" pain relievers; and methamphetamines and "other" stimulants. The final lifetime usage indicators for the parent drugs were created by combining the constituent parts, including the "other" group of substances.

6.3.7.3.1 Hallucinogens

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. It is important to note that if a respondent was a user of at least one of the other hallucinogens, 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. Using the neighborhood created from the hallucinogens' predicted probability of lifetime use, missing values for LSD and/or PCP and/or Ecstasy and/or other hallucinogens were replaced with the values from a donor within this neighborhood. For individuals missing a lifetime usage indicator for either LSD and/or PCP and/or Ecstasy and/or other hallucinogens, only the missing value(s) was (were) replaced. For individuals missing two or more of these lifetime usage indicators, the missing values were replaced by values from the same donor. As with smokeless tobacco, the subcategories for hallucinogens were assigned values separately, making logical constraints unnecessary. As a final step, a lifetime usage indicator for the parent drug was created by combining the lifetime usage indicators for the three subgroups.

---

[49] A respondent could have been asked the gate question for crack only if he or she already indicated use of cocaine.

6.3.7.3.2 Pain Relievers

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 specific 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 this question. 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 questions and "no" responses. Using the neighborhood created from the pain relievers' predicted probability of lifetime use, the missing value(s) for OxyContin and/or other pain relievers was (were) replaced with the value(s) from a donor within this neighborhood. For individuals missing a lifetime usage indicator for either OxyContin or other pain relievers, but not both, only the missing value was replaced. For individuals missing both of these lifetime usage indicators for pain relievers, the missing values were replaced by values from the same donor. As with smokeless tobacco, the subcategories for pain relievers were assigned values separately, making logical constraints unnecessary. As a final step, a lifetime usage indicator for the parent drug was created by combining the lifetime usage indicators for the two subgroups.

6.3.7.3.3 Stimulants

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

6.3.8 Multivariate Imputation for Lifetime Drug Use

Section 6.3.2 summarizes how all of the respondents in the 2005 survey were separated into item respondents and item nonrespondents for the lifetime drug variables. Subsequent sections summarize model building, computation of predicted means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predicted mean. In most cases, however, these univariate assignments were only provisional. As indicated in Table 6.1, the final imputed values for these drug use measures were obtained by building neighborhoods upon a vector of predicted means using the MPMN technique described in Appendix C. In a manner consistent with the univariate imputations, the multivariate assignments were done separately within three age groups: 12 to 17, 18 to 25, and 26 or older. As indicated in earlier sections, a respondent was eligible to have been a donor for a given item nonrespondent if he or she had complete data across all the lifetime drug use variables and was within the same age group.

The values missing for a given respondent define the "pattern of missingness." Respondents with missing lifetime indicators were separated into two groups: respondents missing only one lifetime drug use measure and respondents missing more than one lifetime drug use measure. The respondents missing only one lifetime use indicator were imputed using UPMN. Respondents missing more than one lifetime use indicator were imputed using MPMN.

Only one logical constraint was utilized in the multivariate imputation of lifetime use. Those item nonrespondents who were known to have used pain relievers, but both their OxyContin and "other" pain reliever indicators were missing, were required to have a donor who was a lifetime user of pain relievers. This pattern of nonresponse occurs when respondents respond affirmatively to PR04, but fail to select any drugs from the card in PR04a.

In addition, if possible, donors and recipients were required (as likeness constraints) to come from States with similar drug usage patterns for the drug in question, and donors were required to have each element of the multivariate predictive mean vector "close to" (i.e., within the delta distance of) the recipient's elements of the predictive mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predictive mean vector. The elements of the predictive mean vector corresponded to the predicted values of the recipient's missing lifetime use indicators. Initially, donors and recipients were required to have, if possible, the same values for all nonmissing lifetime use indicators. If this initial constraint did not produce a big enough donor pool, donors and recipients were required to have the same values for only lifetime indicators within the same or related drug modules. The number of respondents for whom donors were found within various likeness constraints is summarized in Appendix G. In general, the likeness constraints were loosened in the following order: (1) remove the requirement that donors and recipients have the same values for all nonmissing lifetime usage indicators; (2) remove the requirement that donors and recipients have the same values for all nonmissing lifetime usage indicators only within a common or related drug module; (3) abandon the neighborhood and choose the donor with the closest predicted mean; and (4) remove the requirement that donors and recipients be from States with similar usage levels.

The full predictive mean vector contained elements for each lifetime drug use measure. However, only a portion of the full predictive mean vector was used. Specifically, only those elements corresponding to the recipient's missing lifetime drug use were used. If the missing lifetime usage indicators corresponded to only one predicted mean, a UPMN imputation similar to the provisional UPMN was utilized. Otherwise, an MPMN imputation was employed. The Mahalanobis distance[50] was then calculated using only the portion of the predictive mean vector associated with the given missingness pattern. If no donors were available who had predicted means within a multivariate delta of the recipient's vector of predicted means, the neighborhood was abandoned and the respondent with the closest Mahalanobis distance was selected as the donor. The procedure is described in detail in Appendix C.

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.

---

[50] See Appendix C for a definition of Mahalanobis distance. A definition also can be found in Manly (1986).

6.4 Editing of Drug Recency of Use, 30-Day Frequency of Use, and Age at First Use

Most of the editing procedures that were applied to the raw data on recency of use, frequency of use, and age at first use are discussed in the 2005 NSDUH editing and coding report (Kroutil et al., 2007). However, a few edits were implemented just before imputation. These 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 date of first use (see Section 6.6.1.8). There are other edits that could have been implemented, but were not. They were not implemented for one of the following reasons:

a) The pattern of inconsistency was not discovered until after processing began.

b) It was decided that the effort required to implement the edit exceeded the benefit derived from this edit.

c) No decision had been made on whether to implement the edit by the time processing began.

6.4.1 Edits Involving "Other" Hallucinogens, "Other" Pain Relievers, and/or "Other" Stimulants

For respondents who were known to have never used "other" hallucinogens, "other" pain relievers, and/or "other" stimulants, certain deductions could be made regarding the relationship between the parent drug data and the child drug data. Note that these edits also could have been applied to respondents who were imputed to lifetime nonuse of the "other" variable.

a) 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 were missing, then the overall hallucinogens recency was assigned to the most recent of the child recencies. (This also was applied for pain relievers and stimulants.)

b) 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.

c) If the respondent was known never to have used "other" hallucinogens (or pain relievers or stimulants), the parent age at first use was nonmissing, only one child age at first use was missing, and the minimum of the nonmissing child ages at first use was greater than the parent age at first use, then the missing child age at first use was edited to the parent age at first use.

6.4.2 Other Age-at-First-Use Edits

a) The following 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 value of the child age-at-first-use value was 3, then the parent age-at-first-use value was edited to 3. This could be deduced because respondents with age-at-first-use values of less than 3 were ineligible to be donors (see Section 6.6.1.6).

b) The following edit applied to all child age-at-first-use variables: daily cigarettes, LSD, PCP, Ecstasy, OxyContin, methamphetamines, and crack. If the parent age at first use was equal to the age, all missing child age-at-first-use values were edited to the age.

c) The following edit also applied to all child age-at-first-use variables. If the parent age at first use was equal to one less than the age, the child recency[51] 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, the child age-at-first-use value was assigned to one less than the age. In particular, the child age at first use must be either less than AGE – 1, greater than AGE – 1, or equal to AGE – 1. It 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 recency implies that the respondent did not use the drug while at his or her current age (since 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. Since the child age at first use cannot be less than AGE – 1 or greater than AGE – 1, it must be equal to AGE – 1.

d) 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: age at first daily cigarette use must have been either less than AGE – 3, greater than AGE – 3, or equal to AGE – 3. The age at first cigarette use precludes the possibility that the age at first daily cigarette use was less than AGE – 3, and the cigarette recency precludes the possibility that the age at first daily cigarette use was greater than AGE – 3.

6.5 Imputation-Revised Drug Recency of Use, 12-Month Frequency of Use, 30-Day Frequency of Use, and 30-Day Binge Drinking Frequency

In the 2005 survey, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and (for alcohol) 30-day binge drinking frequency[52] were modeled separately for each drug. These measures of drug usage constituted a multivariate set within each drug. Provisional values replaced missing values for use in subsequent models, where necessary, using the UPMN methodology, as described in Appendix C. After having modeled all of the drug use measures for a given drug, the MPMN methodology (also described in Appendix C)

---

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

[52] "Binge drinking" was defined as having five or more drinks on the same occasion on a given day. The 30-day binge drinking frequency was defined as the number of days out of the past 30 where the respondent had five or more drinks on the same occasion.

was employed to determine final imputed values using the predicted values from these models. Separate multivariate imputations were conducted for each drug.

The implementation of the PMN methodology required the identification of a modeling hierarchy, as described in Appendix C. However, for the multivariate imputations described in this section, two separate modeling hierarchies were employed. Within a multivariate set, recency of use was modeled first, followed by the 12-month frequency of use (where applicable), 30-day frequency of use (where applicable), and (for alcohol) 30-day binge drinking frequency. Once the multivariate imputation for a given drug was completed, the recency of use for the next drug in the sequence was modeled.

6.5.1 Recency of Use

6.5.1.1 Hierarchy of Drugs

A complete drug hierarchy, as described in Appendix C, was not required for recency of use because only cigarettes, alcohol, and marijuana recencies were used as covariates in models for subsequent drugs. This was due to difficulties that would have arisen if too many covariates were included in the polytomous logistic models. (Lifetime usage indicators of other drugs were included instead of recency-of-use indicators.) However, for the sake of convenience, the recency of use imputations did follow the same hierarchy as described in Section 6.2.

6.5.1.2 Setup for Model Building and Hot-Deck Assignment

As with all the drug use measures, the recency-of-use imputations were conducted separately for 12- to 17-year-olds, 18- to 25-year-olds, and respondents aged 26 or older. To impute missing recency-of-use values for each drug, it was first necessary to define the eligible population within each of these age groups. Using the imputation-revised lifetime indication of use, the file was reduced to lifetime users. Among these lifetime users, item respondents and nonrespondents for each drug were identified across recency of use and (where applicable) the 12-month, 30-day, and (for alcohol only) 30-day binge drinking frequency-of-use measures. If a valid response was provided for each drug use measure, the person was deemed an item respondent for the drug. Otherwise, he or she was an item nonrespondent.

Before modeling, the respondents' weights were adjusted so that they represented all lifetime users. (Weights were defined in the same way as with other drug use variables. See discussion about how the weights were defined in Section 6.3.2.) Because item respondents were defined at the drug level, these adjustments were made separately for each drug (and within the three age groups). The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model included imputation-revised cigarette, alcohol, and marijuana recencies (where applicable); lifetime indicators of usage of drugs other than cigarettes, alcohol, and marijuana; gender; age;[53] race/ethnicity; first-order interaction of gender and race/ethnicity; marital status; education level;

[53] The covariate "categorical age" was divided into five categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). For the 12-to-17 and 18-to-25 age groups, categorical age was not included as a covariate in the item response propensity models.

employment status;[54] census region; and an MSA[55] indicator.[56] In addition, a three-level State rank variable was defined by clustering States according to the prevalence of past month use of the drug of interest and was included as a covariate in the models.[57]

6.5.1.3 Sequential Model Building

Using the adjusted weights, the probability of selecting each recency-of-use category was modeled within each age group using, where possible, polytomous logistic regression.[58] The predictors included in the models were imputation-revised cigarette, alcohol, and marijuana recencies (where applicable); lifetime indicators of usage of drugs other than cigarettes, alcohol, and marijuana; centered age;[59] centered age squared; centered age cubed; gender; race/ethnicity; first-order interactions of centered age, gender, and race/ethnicity; marital status; education level; employment status;[60] census region; an MSA indicator; and State rank.[61] Because interest was only in the estimation of the predicted mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. For a summary of the variables included in each drug model, see Appendix F.

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[62] for these drugs caused convergence problems when fitting multinomial 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

---

[54] Marital status, education level, and employment status were included as covariates for the 18-to-25 and 26-or-older age groups only.

[55] Metropolitan statistical area, as defined by the Office of Management and Budget (OMB).

[56] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

[57] For drug/age group combinations where the proportion of past month users was low, age groups were aggregated within the given drug. Sometimes all three age groups were aggregated, and other times two neighboring age groups were aggregated. This was done mainly to keep the State rank from being easily influenced by only one or two users in the sample. Also, all States with no users were placed in the lowest State rank category, so for especially rare drugs such as heroin and sedatives, substantially more than one third of the States received the lowest State rank.

[58] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in RTI (2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

[59] The covariate age was centered within each age group in order to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[60] Marital status, education level, and employment status were included as covariates for the 18-to-25 and 26-or-older age groups only.

[61] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

[62] A "lopsided distribution" in the context of recency of use is where, among the categories 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."

problem, the single multinomial logistic model was replaced with two binary logistic models[63] that were fitted in a hierarchical manner. The following paragraphs describe the two models.

As with the multinomial logistic model, the first binary logistic model was fitted 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. (Weights were defined in the same way as with other drug use variables. See the discussion about weights in Section 6.3.2.) 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 utilizing the predicted means from the first model.)

From the two binary logistic models, both the probability of past month use and the probability of past year but not past month use were obtained and utilized in the provisional hot deck program for recency, which is discussed in subsequent sections. Once the predicted means were determined from the two models, a single vector of predicted means conditional on lifetime usage, as with the multinomial logistic models, was determined in the following manner:

1. *P*(past month use | lifetime use) =

   P(past month use | past year use)**P*(past year use | lifetime use)

2. *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).

6.5.1.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

Because recency-of-use and the frequency-of-use variables for a given drug were considered part of a multivariate set, the calculation of predicted means for the frequency-of-use variables required the item nonrespondents to have been identified as provisional past month and/or past year users. Within a given drug and within each age group, predicted probabilities for each of the recency categories were computed for both item respondents and item nonrespondents using the parameters from the appropriate logistic model(s). The predicted probabilities from the recency models were used to assign provisional values using the UPMN imputation method described in Appendix C. A vector of predicted probabilities for each respondent was created by the logistic regression model(s). Because only a single predicted mean was used to determine the neighborhood when determining provisional values, not all of the

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

predicted probabilities from the model were used.[64] Also, because past month use was the most critical measure of recency of drug use, the neighborhoods were defined based on the probability of past month use. If possible, provisional donors were chosen with predicted means within the delta of the recipient, where the value of delta varied depending on the value of the predicted means, which in this case were predicted probabilities of past month use.[65] In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5 and was defined as 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5 and allowed a tighter delta for predicted probabilities close to 0 or 1. If no donors were available with predicted means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predicted mean was chosen.

6.5.1.5 Assignment of Provisional Imputed Values

Subject to the constraints described in the next section, separate assignments of provisional values were performed within each of the three age groups. The final recency-of-use imputations were multivariate across drug measures and are further described in Section 6.5.5.

6.5.1.6 Constraints on Univariate Predictive Mean Neighborhoods

As stated in the lifetime usage section, a UPMN neighborhood can be restricted by logical constraints (which cannot be loosened) and by likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible. The likeness constraints and logical constraints, which were applied, are described separately in the next two paragraphs.

As with all other drug use measures, neighborhoods for recency of use were restricted so that candidate donors and recipients would have been within the same age group (12 to 17, 18 to 25, or 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. A small delta also could have been considered a likeness constraint, which could have been loosened by enlarging or removing delta. As previously stated, if no donors were found in the delta, as defined in Section 6.5.1.4, the neighborhood was then abandoned and the donor with the predicted mean closest to the recipient was chosen.[66] If possible, donors and recipients were required to be from States with the same level of usage of a given drug (the State rank, as defined in the introduction of this chapter), where the level of usage was defined in terms of the proportion of a given State's residents who had used a given drug in the past month. If insufficient donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned and the donor with the

[64] A multivariate procedure could have been used to determine the provisional values that would have been used for all of the predicted probabilities in the predictive mean vector. However, the amount of effort and computation time associated with multivariate imputation is considerably greater with multivariate procedures than with univariate procedures. Because the imputation was only provisional, a univariate imputation was used.

[65] The probability of past month use was used to define univariate neighborhoods for recency of use even when it was known that the respondent was not a past month user. This could occur if the edited recency of use was, for example, lifetime not past month use.

[66] Although using neighborhoods is important for the calculation of the variance due to imputation, methods to account for donor-predicted means differing greatly from recipient-predicted means had not yet been devised at the time these imputations were implemented.

closest predicted mean was chosen; and (2) donors and recipients were no longer required to be from States with similar usage levels.

The only logical constraints placed on the neighborhoods involved cases where a general recency category was available for a respondent and imputation was required to determine the specific recency categories. The general recency categories that appeared are shown in Table 6.5. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood. Other logical constraints involving a very small number of respondents were not applied to the provisional imputations. The complete list of constraints used in the multivariate imputation of recency and frequency of use is provided in Section 6.5.5.

Table 6.5 General Incomplete Recency Categories for Tobacco and Nontobacco

General Incomplete Recency Category	Allowable Specific Recency Categories (Tobacco)	Allowable Specific Recency Categories (Nontobacco)
Lifetime	1. Lifetime but not past 3 years 2. Past 3 years but not past year 3. Past year but not past month 4. Past month	1. Lifetime but not past year 2. Past year but not past month 3. Past month
Past Year	1. Past year but not past month 2. Past month	1. Past year but not past month 2. Past month
Lifetime, Not Past Year	1. Lifetime but not past 3 years 2. Past 3 years but not past year	N/A (for nontobacco, this was a specific recency category)
Lifetime, Not Past Month	1. Lifetime but not past 3 years 2. Past 3 years but not past year 3. Past year but not past month	N/A
Lifetime, Not Past Month, but Within Past Three Years[1]	1. Past 3 years but not past year 2. Past year but not past month	N/A
Past Three Years[1]	1. Past 3 years but not past year 2. Past year but not past month 3. Past month	N/A

[1]These incomplete recency categories were created for the first time in the 2005 NSDUH.

6.5.1.7 Multivariate Assignments

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 methamphetamines (child); and hallucinogens (parent) and LSD, PCP, and Ecstasy (children). The multivariate assignment of imputed values for chew and snuff is discussed below.

For reasons discussed in Section 6.3.7.1, 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. At this stage in the process, lifetime use or nonuse of either chew or snuff was considered known (employing information from the lifetime usage imputation). For lifetime users of chew or snuff who were missing some or all of their recency-of-use information for either chew or snuff, but not both, only the missing specific recency-of-use values were replaced.[67] However, for individuals missing recency-of-use information for both chew and snuff (given that the respondent was known or was imputed to have been a chew user and a snuff user), values for both were obtained from the same donor. The provisional recency of use for smokeless tobacco was obtained by combining the recency-of-use information from chew and snuff.

6.5.2 12-Month Frequency of Use

6.5.2.1 Hierarchy of Drugs

The modeling of 12-month frequency sequentially followed that of recency of use for each drug. Across drugs, the sequence was exactly the same as the one used for recency of use. Data on 12-month frequency of use were not collected for all of the drugs. Thus, these imputations were conducted for a subset of the drugs (see Table 6.1).

6.5.2.2 Setup for Model Building and Hot-Deck Assignment

As with all the drug use measures, the 12-month frequency-of-use imputations were conducted separately for 12- to 17-year-olds, 18- to 25-year-olds, and respondents aged 26 or older. 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). Among the past year users of each drug, item respondents, item nonrespondents, and the response propensity adjustment were defined. 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. (Weights were defined in the same way as with other drug use variables. See discussion about how the weights were defined in Section 6.3.2.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The predictors in the response propensity adjustment modeling included the provisional indicator of past month use for the drug of interest; (where available) recencies of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers,

---

[67] For respondents missing all of their recency information, the only known information was that they were lifetime users (either from their survey response or from imputation). For respondents missing some of their recency information, they might have been assigned a general recency category (outlined in Table 6.3), and if so, then specific recency values were imputed.

tranquilizers, stimulants, and sedatives;[68] categorical age;[69] race/ethnicity; gender; census region; and an MSA indicator.[70]

6.5.2.3 Model Building

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 have been 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 at first use and the date of the interview indicated the total time period during which they could have been using drugs.[71] 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 have been 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. 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.

---

[68] 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 (past month use versus past year but not past month use) was always defined.

[69] The covariate "categorical age" was divided into five categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). For the 12-to-17 and 18-to-25 age groups, categorical age was not included as a covariate in the item response propensity models.

[70] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

[71] 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, the maximum period the respondent could have been using drugs was assumed to be 365.25 because no other information was available.

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.[72] 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.[73] Thus, the models included: a provisionally imputed indicator of past month use of the given drug; (where available) the imputation-revised recencies of use for cigarettes,[74] smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; centered age;[75] centered age squared; centered age cubed; gender; race/ethnicity; first-order interactions of centered age, gender, and race/ethnicity; marital status; education level; employment status;[76] census region; an MSA indicator; and State rank (based on past month prevalence of the drug).[77] Because interest focused only on the estimation of the predicted mean and not on the parameter estimates exclusively or their standard errors, no model selection was attempted. Predicted 12-month frequencies of use were defined by back-transforming the resulting predicted values. For a complete summary of the 12-month frequency-of-use models, see Appendix F.

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 transformed back into a proportion for use as the variable from which the neighborhoods were created. This proportion could have been treated as a probability, which, in turn, could have been 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.

---

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

[73] As with the recency-of-use models, for a handful of cases, the State rank variable could not have been 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.

[74] The covariates based on recency-of-use variables that corresponded to drugs other than the one being modeled (if the recency of use was available) were defined by a series of dummy variables reflecting the different recency categories.

[75] The covariate age was centered within each age group in order to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[76] Marital status, education level, and employment status were included as covariates for the 18-to-25 and 26-or-older age groups only.

[77] These variables were included in every model, unless small samples sizes precluded the use of such a large pool of covariates. If this occurred, the model was reduced.

6.5.2.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

Within a given drug, predicted means from the 12-month frequency-of-use models were computed for both item respondents and item nonrespondents using the parameters from the regression model. The logits were converted back to proportions, which were, in turn, multiplied by the probability of past year use to make the predicted mean conditional on lifetime use.[78] Using the UPMN methodology described in Appendix C, neighborhoods were defined based on these predicted means. If possible, provisional donors were chosen with predicted means within delta of the recipient, where the value of delta varied depending on the value of the predicted means, which in this case were predicted proportions of the year used. In particular, delta was defined as 5 percent of the predicted proportion if the proportion was less than 0.5 and was defined as 5 percent of 1 minus the predicted proportion if it was greater than 0.5. This allowed a looser delta for predicted proportions close to 0.5 and allowed a tighter delta for predicted proportions close to 0 or 1. As with recency of use, if no donors were available with predicted means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predicted mean was chosen.[79]

6.5.2.5 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. Separate assignments were performed within each of the three age groups, subject to the constraints described in the next section. For the 12-month frequency of use, "level of usage" for the State rank groups was defined in terms of the proportion of a given State's residents who had used a given drug in the past month. Assignments were not required for tobacco because the tobacco module did not have 12-month frequency-of-use questions. Also, assignments were not needed for "pills"[80] because pills did not have a 30-day frequency-of-use question, making it unnecessary to obtain provisionally imputed 12-month frequencies. The final 12-month frequency-of-use imputations were multivariate across drug measures and are further described in Section 6.5.5.

---

[78] The dependent variable in the model used was the empirical logit, as described in Section 6.5.2.3. The back-transformed value was obtained by solving for *Y/N*, where *Y* is the number of days of use (in a year) and *N* is the number of potential days of use in the year.

[79] Although using neighborhoods is important for the calculation of the variance due to imputation, methods to account for donor-predicted means differing greatly from recipient-predicted means had not yet been devised at the time these imputations were implemented.

[80] "Pills" were defined as pain relievers, tranquilizers, stimulants, and sedatives.

6.5.2.6 Constraints on Univariate Predictive Mean Neighborhoods

An obvious logical constraint for 12-month frequency of use was that all donors were past year users. Other logical constraints involved the interview date, month of first use, birthday, recency of use, and 30-day frequency of use. See Section 6.5.5 for a discussion of the multivariate imputation of recency and frequency of use.

Two likeness constraints used in the assignment of values for 12-month frequency of use were identical to those of recency of use: the three age groups and the State rank groups based on level of past month usage. As with the recency-of-use models, delta was set so that the predicted means of all potential donors were within 5 percent of the item nonrespondent's predicted mean, where the predicted mean was defined to be the proportion of the year (or maximum period within a year) during which a respondent used a drug. Finally, recipients and donors were required to have the same recency of use (past month versus past year but not past month), whether that recency of use was reported or imputed.[81] If no donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned and the donor with the closest predicted mean was chosen; (2) donors and recipients were no longer required to be from States with similar usage levels; and (3) donors and recipients were no longer required to have the same recency of use.

Occasionally, more than one substance was associated with a single predicted mean. However, for the provisional imputed values, only the "parent" drug was of interest (for example, only the provisionally imputed cocaine 12-month frequency was needed and not the crack 12-month frequency). Therefore, multivariate assignments were not needed for the provisional UPMNs, but they did occur in the final multivariate imputation of recency and frequency.

6.5.2.7 Multivariate Assignments

Although more than one substance was occasionally associated with a single predicted mean, the provisionally imputed 12-month frequencies were required only if they were needed for calculating predicted means using the coefficients from a subsequent model. 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. However, no 12-month frequency was asked of smokeless tobacco users. Moreover, no provisionally imputed values were required for substances that were a subset of the substance associated with the predicted mean, which have been referred to as "parent/child" drugs (see Section 6.2). Hence, no multivariate assignments were required for the provisionally imputed 12-month frequency.

[81] 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.

6.5.3 30-Day Frequency of Use

6.5.3.1 Hierarchy of Drugs

The modeling of 30-day frequency followed that of recency and 12-month frequency of use for each drug. Across drugs, the sequence was exactly the same as that for recency of use. Data on 30-day frequency of use were not collected for all of the drugs. Thus, these imputations were performed for only a subset of the drugs (see Table 6.1).

6.5.3.2 Setup for Model Building and (for Alcohol Only) Hot-Deck Assignment

The file was first reduced to the eligible population, which was 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 have been 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. (Weights were defined in the same way as with other drug use variables. See the discussion in Section 6.3.2 about how the weights were defined.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. Predictors for the response propensity models included: the provisional 12-month frequency for the drug of interest (where applicable); (where available) recencies of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives;[82] categorical age;[83] race/ethnicity; gender; census region; and an MSA indicator.[84]

6.5.3.3 Model Building

As is apparent from the previous section, only past month users of the drug of interest were used to build the 30-day frequency-of-use model. The response variable of interest in the 30-day frequency-of-use models for most drugs, prior to a normalizing transformation, was the proportion of the days in a month (30) on which a respondent used a particular drug. 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:

$$\log\left[(Y_i + 0.5)/(N - Y_i + 0.5)\right],$$

[82] 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, since all respondents in the model were past month users.

[83] The covariate "categorical age" was divided into five categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). For the 12-to-17 and 18-to-25 age groups, categorical age was not included as a covariate in the item response propensity models.

[84] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

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:

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

where P_i was defined as the proportion of days in the past year on which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[85] 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.

Because the 30-day frequency models were limited to past month users, only one provisional recency category was relevant for the drug of interest.[86] Hence, provisional recency of use for the drug of interest could not have been included in the 30-day frequency-of-use model. However, imputation-revised recency of use of other drugs could have been included. For drugs where the recency of use was not yet modeled, the lifetime indication of use served as a surrogate for the recency-of-use indicators. Covariates representing the State rank groups (defined by the level of past month use) were included to adjust for any State drug use differences. Other covariates included: the provisional 12-month frequency of use for the drug of interest (where applicable); census region; centered age;[87] centered age squared; centered age cubed; gender; race/ethnicity; the first-order interactions of centered age, gender, and race/ethnicity; marital status; education level; employment status;[88] and an MSA indicator.[89] Because interest was only in the estimation of the predicted mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. The predicted 30-day frequencies of use were defined by back-transforming the predicted values from the models. For a complete summary of the 30-day frequency-of-use models, see Appendix F.

The predicted mean that came out of the 30-day frequency-of-use model was a logit of the proportion of the month used. This logit was transformed back into a proportion for use as the variable from which the neighborhoods were created. This proportion was treated as a probability, which, in turn, was multiplied by the probability of past month use in order to have made the predicted means conditional on lifetime use of the drug in question.[90] When calculating predicted means for some item nonrespondents, sometimes it was not known whether they were

---

[85] 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.)

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

[87] The covariate age was centered within each age group in order to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[88] Marital status, education level, and employment status were included as covariates for the 18-to-25 and 26-or-older age groups only.

[89] These variables were included in every model unless small samples sizes precluded the use of such a large pool of covariates. If this occurred, the model was reduced.

[90] The dependent variable in the model used was the empirical logit given in Section 6.5.3.3. The back-transformed value was obtained by solving for Y/N, where Y is the number of days of use (in a month) and N is the number of potential days of use in the month (30.4375).

past month users or not. 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.

For cigarettes, chewing tobacco, and snuff, 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. These substances were 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), a model much like the 30-day frequency-of-use models for other substances 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. The same pool of covariates was used in the logistic model and the regression model with the logit as the response variable. It should be noted that, unlike recency of use, the 30-day frequencies for chewing tobacco and snuff were not combined into a single value for smokeless tobacco. Since it was not possible to determine if the x days using chewing tobacco overlapped with the y days using snuff, separate models were fitted for chewing tobacco and snuff.

6.5.3.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

Within a given drug, predicted means from the 30-day frequency-of-use models were computed for both item respondents and item nonrespondents using the parameters from the regression model. The 30-day frequency models were fitted after recency of use and 12-month frequency of use. 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 the UPMN technique described in Appendix C. 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).

6.5.3.5 Assignment of Provisional Imputed Values (Alcohol Only)

Separate assignments for the 30-day frequency of alcohol use were performed within each of the three age groups, subject to the constraints described in the next section. For the 30-day frequency of use, "level of usage" was defined in the same manner as the recency of use and 12-month frequency of use.

6.5.3.6 Constraints on Univariate Predictive Mean Neighborhoods (Alcohol Only)

An obvious logical constraint was that all donors had to have been past month users. In addition, the donated 30-day frequency was required to have been less than or equal to the respondent's preexisting 12-month frequency—whether that 12-month frequency was reported or imputed—and greater than or equal to the respondent's preexisting 30-day binge drinking frequency. Two likeness constraints used in the assignment of values for 30-day frequency of use were identical to those used for recency of use and 12-month frequency of use. The two likeness constraints were the three age groups and the State rank groups based on level of past month usage. As with the recency-of-use models, delta was set so that the predicted means of all potential donors were within 5 percent of the item nonrespondent's predicted mean, where the predicted mean was defined to have been the proportion of the month during which a respondent used a drug. If no donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned and the donor with the closest predicted mean was chosen; and (2) donors and recipients were no longer required to be from States with similar usage levels.

6.5.3.7 Multivariate Assignments

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 for calculating predicted means using the coefficients from 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. Since alcohol is not a "parent/child" drug (see Section 6.2 for a definition of "parent/child" drug), no multivariate assignments were required for provisionally imputed 30-day frequency.

6.5.4 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. However, the 30-day binge drinking frequency model included the provisional alcohol 30-day frequency of use[91] as a covariate. Moreover, the 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.

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 drank five or more drinks. The range of values for the proportion was from 0 to 1. 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:

[91] The provisional 30-day frequency of use was defined by randomly selecting donors from univariate neighborhoods, which were defined by using the respondent and nonrespondent predicted values.

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

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.[92] 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 MPMN imputation. No UPMN step was taken, and no provisional imputed values were determined.

6.5.5 Multivariate Imputation for Recency of Use, 12-Month Frequency of Use, 30-Day Frequency of Use, and 30-Day Binge Drinking Frequency

Sections 6.5.1, 6.5.2, 6.5.3, and 6.5.4 summarize how the set of lifetime drug users in the sample of the 2005 survey was separated into item respondents and item nonrespondents for the recency of use, 12-month frequency of use, 30-day frequency of use, and (for alcohol) 30-day binge drinking frequency drug use measures. These sections also summarize model building, computation of predicted means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predicted mean. In most cases, however, these univariate assignments were only provisional. As is indicated in Table 6.1, the final imputed values for these drug use measures were obtained by building neighborhoods upon a vector of predicted means using the MPMN technique described in Appendix C. In a manner consistent with the univariate imputations, the multivariate assignments were done separately within three age groups: 12 to 17, 18 to 25, and 26 or older. As indicated in earlier sections, a respondent was eligible to have been a donor for a given item nonrespondent if he or she had complete data across the drug use measures for the drug in question and was within the same age group. As with the provisional imputations, the donated value for the 12-month frequency of use variable was determined by taking the product of the donated proportion of the year that the donor had used the substance of interest and the recipient's maximum number of possible days that he or she could have used the substance.

6.5.5.1 Constraints on Multivariate Predictive Mean Neighborhoods

6.5.5.1.1 Logical Constraints

The logical constraints required in the provisional univariate imputations discussed in Sections 6.5.1, 6.5.2, and 6.5.3 also were required in the multivariate imputations. However, some constraints that potentially could have been applied in the provisional recency-of-use and provisional 12-month frequency imputations were not applied because of the very small number of respondents affected, and thus they are not listed in Table 6.5 or mentioned in Sections 6.5.1 or 6.5.2. However, these constraints were applied in the multivariate imputations. In particular,

[92] 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 0, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[0.5/(N + 0.5)]$. (See Cox and Snell [1989] for a discussion of the empirical logit transformation.)

the possible recencies of use were limited based on the respondent's current age, the time between the interview date and the birthday, the time between the interview date and the month of first use, and any nonmissing frequency-of-use information. In addition, if the respondent was (or could have been) a past month user and was known to have used the drug at least once between 1 month before the interview date and 1 year before the interview date (because of the given month and/or year of first use), donors were required to have a 12-month frequency that reflected this. In general, the application of these constraints depended on what information was missing in the recency-of-use and frequency-of-use variables. The values missing for a given respondent define the "pattern of missingness." For example, one pattern of missingness for marijuana could be as follows: past year user of marijuana (recency partially missing), 12-month frequency not missing, and 30-day frequency missing. In this example, the logical constraints have to make the imputed 30-day frequency consistent with the preexisting 12-month frequency. In the case where the 12-month frequency of use variable was missing, an additional logical constraint involved the product of the donated proportion and the recipient's maximum possible number of days used in a year (called the "donated 12-month frequency product"). Since this product involved both the donor and the recipient, it had to be consistent with the 30-day frequency of use, regardless of whether the 30-day frequency was a preexisting nonmissing value or a donated value. It also had to be greater than 1 and/or greater than the 30-day frequency when it was known that the respondent was a past month user, but started using prior to the past month in the past year. The various patterns of missingness for each drug, the logical constraints imposed on the set of donors, and the frequency with which each missingness pattern occurred are provided in Appendix H.

6.5.5.1.2 Likeness Constraints

In addition, if possible, donors and recipients were required (as likeness constraints) to come from States with similar drug usage patterns for the drug in question, and donors were required to have each element of the multivariate predictive mean vector "close to" (i.e., within the delta distance) the recipient's elements of the predictive mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predictive mean vector. Finally, for drug modules with multiple substances (i.e., parent/child relationships), if the recency of use for one or more of the substances within the module was not missing, donors and recipients were required to have, if possible, the same values for these recency-of-use indicators.[93] The number of respondents for whom donors were found within various likeness constraints is summarized in Appendix G. In general, the likeness constraints were loosened in the following order: (1) for drug modules with multiple substances, likeness constraints requiring donors and recipients to have had the same recency-of-use values for nonmissing variables were removed, while any necessary logical constraints were maintained; (2) the neighborhood was abandoned and the donor with the closest predicted mean was chosen; and (3) donors and recipients were no longer required to be from States with similar usage levels.

[93] Donors also were required to match the recipient with respect to lifetime use of "other" drugs for hallucinogens. During the processing, this constraint was loosened later than the constraint involving the recency-of-use indicators for the three child drugs. No likeness constraints involving "other" drugs were applied to pain relievers or stimulants.

6.5.5.1.3 More Than One Substance for a Single Predictive Mean Vector

Occasionally, more than one substance was associated with a single predicted mean, whether it was for recency of use or the frequency-of-use variables. This could have been two substances of equal standing considered together when modeling (snuff and chewing tobacco) or drugs with a parent/child relationship (see Section 6.2 for a definition of parent/child relationship). The assignment of imputed values for these substances was unique for each situation. Hence, the imputations for each of these substances are discussed as follows.

Smokeless Tobacco. As noted in Sections 6.3.7.1 and 6.5.1.7, one model for smokeless tobacco recency of use (a combination of the chew and snuff responses) was fitted rather than individual models for chew and snuff. The nearest neighbor hot-deck neighborhood was then based on the predicted probability of past month use of smokeless tobacco. The assignment of recency-of-use values for smokeless tobacco followed the same logical constraints in the multivariate imputation as was the case for the univariate imputations discussed in Section 6.5.1.7.

Unlike recency of use, however, separate models for chew and snuff were built for 30-day frequency of use. The predicted means from these models were conditioned on past month use. In the 30-day frequency-of-use imputations, discussed in Section 6.5.3.3, the predicted means used to form the neighborhoods were conditioned on lifetime usage rather than past month usage. Because the 30-day frequency models gave predicted means conditioned on past month use, it was necessary to determine the probability of past month use given lifetime use, which could have been obtained from the recency models. Because the 30-day frequencies for chew and snuff could not have been combined, recency-of-use models were built for chewing tobacco and snuff separately.[94] (This was in addition to the regular recency-of-use model that was built for smokeless tobacco.) The covariates used in the models are provided in Appendix F.

Cocaine and Crack. Even though cocaine and crack were in distinct modules, single models were fitted for recency of use and the frequency-of-use variables using the information from the cocaine module. Crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. As with smokeless tobacco, use or nonuse of crack was considered known (using information from the lifetime imputations). Hence, as a logical constraint, users of crack with incomplete recency (or frequency) information required donors who were also crack users. Moreover, if the cocaine recency was not missing, the donated crack recency could not have been more recent than the preexisting cocaine recency. Similarly, if the crack recency was not missing, but the cocaine recency was missing, the donated cocaine recency could not have been less recent that the preexisting crack recency.

If at least one of the frequency-of-use variables was missing, but the cocaine recency was not, the cocaine recency of use for donors and recipients had to match. In addition, donors and recipients were required to have the same crack recency of use if it was known that the recipient

---

[94] To properly condition the respective 30-day frequency predicted means for chewing tobacco and snuff, it was not possible to use the predicted probabilities available for the recency of use of smokeless tobacco as a whole. Instead, separate recency-of-use models for chewing tobacco and snuff were used to obtain the predicted probabilities of both past month use and past year but not past month use of these substances. These were the values utilized in the construction of conditional probabilities for the 30-day frequencies of chewing tobacco and snuff. See Appendix H for details.

used crack in the past year. Both of these constraints were applied regardless of the pattern of missingness among the frequency-of-use variables. Additional logical constraints involved the "donated 12-month frequency products" for both crack and cocaine. If both the crack and cocaine 12-month frequency of use values were missing, it was necessary to check the donated products against each other for consistency since this product depended upon both the donor and recipient, even though the donated proportions came from the same donor. Both also had to be checked for consistency against the 30-day frequency-of-use values (if the respondent was a past month user of crack and/or cocaine), regardless of whether those variables were preexisting nonmissing values or donated imputed values. If only one of the 12-month frequency-of-use variables were missing, the donated product was checked for consistency against the preexisting nonmissing 12-month frequency of use value and against the 30-day frequency of use variables, imputed or not.

Hallucinogens (LSD, PCP, Ecstasy, and Other Hallucinogens), Pain Relievers (OxyContin and Other Pain Relievers), and Stimulants (Methamphetamines and Other Stimulants)

As stated in Section 6.3.7.3, the modules for hallucinogens, pain relievers, and stimulants included subgate questions referring to child drugs. Hallucinogens had three child drugs (LSD, PCP, and Ecstasy); pain relievers had one (OxyContin); and stimulants had one (methamphetamines). Recency-of-use information for the parent drugs was used in subsequent models, and recency-of-use information for the child drugs was not used. Hence, obtaining provisional values for the recency of use of the child drugs was not necessary. Predictive recency probabilities were calculated for the parent drugs, and these probabilities were used to determine neighborhoods for each group of drugs. As with smokeless tobacco, use or nonuse of the child drugs was considered known (including values that were imputed in the lifetime usage imputations).

Hallucinogens. Using the neighborhood created from the predictive mean vector, missing specific recency categories for LSD and/or PCP and/or Ecstasy and/or hallucinogens as a whole, were replaced with the specific recency categories from a single donor. Child drug (LSD, PCP, and/or Ecstasy) users with incomplete recency information were constrained to have donors who were lifetime users of the specific child drug(s). Moreover, donors were constrained so that a preexisting child drug recency could not have been more recent than a donated parent drug recency. Conversely, a preexisting parent drug recency value could not have been less recent than any donated child recency value. In addition, donors were constrained for those respondents missing the parent recency who used no "other" type of hallucinogen so that the donated parent recency was equal to the minimum of the child recencies, whether donated or not. For individuals missing recency information for the parent drug or the child drugs, only the missing value(s) was (were) replaced. For individuals missing recency information for two or more of these substances, the missing categories were replaced by values from the same donor.

No 12-month frequency-of-use variables were available for any of the three child drugs. However, the "donated 12-month frequency product" for all hallucinogens was required to have been consistent with the 30-day frequency-of-use value for all hallucinogens, whether it was imputed or was a preexisting nonmissing value.

Pain Relievers. A similar procedure was followed for the pain relievers module. Using the neighborhood created from the predictive mean vector, missing specific recency-of-use categories for OxyContin and/or pain relievers as a whole, were replaced with the specific recency categories from a single donor within this neighborhood. OxyContin users with incomplete recency information were constrained to have donors who were also OxyContin users. Moreover, donors were constrained so that a preexisting OxyContin recency-of-use value could not have been more recent than a donated pain relievers recency-of-use value, and, conversely, a preexisting pain reliever recency-of-use value could not have been less recent than the donated OxyContin recency of use. In addition, donors were required to have an overall pain reliever recency equal to their OxyContin recency for those respondents missing both overall pain reliever recency and OxyContin recency, who used no "other" type of pain reliever. For individuals missing recency information for OxyContin and/or pain relievers, as a whole, only the missing categories were replaced. For individuals missing recency information on both of these substances, the missing categories were replaced by values from the same donor.

The major difference between hallucinogens and pain relievers was that a 12-month frequency-of-use variable was available for the child drug, OxyContin. Even though separate 12-month frequency questions were asked for overall pain relievers and more specifically for OxyContin, 12-month frequency was modeled for overall pain relievers only. As with cocaine and crack, additional logical constraints involved the product of the donated proportion and the recipient's maximum possible number of days used in a year (called the "donated 12-month frequency product") for both OxyContin and pain relievers. If both the pain relievers and OxyContin 12-month frequency-of-use values were missing, it was necessary to check the donated products against each other for consistency since this product depended upon both the donor and recipient, even though the donated proportions came from the same donor. No additional check was necessary since pain relievers did not have a 30-day frequency-of-use variable. If only one of the 12-month frequency-of-use variables was missing, the donated product naturally was checked for consistency against the preexisting nonmissing 12-month frequency-of-use value.

Stimulants. The procedure used for the stimulants module was very similar to the one followed for the pain relievers module. As for the pain relievers, a 12-month frequency-of-use variable was available for the child drug (methamphetamines). The constraints that were applied, and the predicted means that were used, were the same as for pain relievers.

6.5.5.2 Final Multivariate Assignment

The full predictive mean vector contained several elements for recency of use (different probabilities associated with each of the recency categories), as well as elements for the frequency-of-use variables. 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 resulting elements in the predictive mean vector that could have potentially resulted are shown in Table 6.6. It is important to note that not all drugs contained all the elements given, as is apparent by looking at the rightmost column in Table 6.6. It should be noted that Table 6.6 assumes 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. Table 6.7 shows the full predictive mean vector for each drug. 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 to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are provided in Appendix H. 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. If no donors were available who had predicted means within a multivariate delta of the recipient's vector of predicted means, the neighborhood was abandoned and the respondent with the closest Mahalanobis distance was selected as the donor. The procedure is described in detail in Appendix C.

Table 6.6 Elements of Full Predictive Mean Vector

Drug Use Measure and Category of Interest	Predicted Mean	Substances
Recency of Use, Past Month[1]	*P*(past month user | lifetime user)	All substances
Recency of Use, Past Year But Not Past Month[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[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 (exc. Alcohol)	*P*(use on a given day in the month | past month user, not a daily user)**P*(not a daily user | lifetime user)**P*(past month user | lifetime user)[5]	Cigarettes, chewing tobacco, snuff
Daily Use	*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] Note that 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 added to 1, and this probability was determined by the other probabilities.

[2] "Tobacco products" included: 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. This, of course, was not true. However, 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 footnote #3 within this table).

[6] "Chew" was short for "chewing tobacco."

[7] "Pills" included pain relievers, tranquilizers, stimulants, and sedatives.

The construction of the predictive mean vectors for the drug families mentioned in Section 6.5.5.1.3 was often complex. The main reason for the complexity is that recency and frequency models were not fit for all child drugs. The predicted means from the models for the parent drug were often used as surrogates for the child drug predicted means. When constructing the predictive mean vectors, the following general principles were followed:

a) If both the parent drug recency and the child drug recency/ies were missing, condition on the general recency category of the parent drug.

b) For smokeless tobacco, if both the chewing tobacco recency and the snuff recency were missing, condition on whichever was "more" missing. Specifically, condition the recency predictive mean vectors on the more general recency category. For example, if chewing tobacco recency was "not past month" and snuff recency was "not past year", condition on the chewing tobacco recency category, since it was more general.

c) Condition all elements of the predictive mean vector on the same general recency level.

Table 6.7 Full Predictive Mean Vector for Sample Drugs

Drug Use Measure and Category of Interest	Drug			
	Tobacco Products[1]	Alcohol	Marijuana, Cocaine, Crack, Heroin, Inhalants, Hallucinogens	Pain Relievers, Stimulants, Sedatives, Tranquilizers
Recency of Use, Past Month Use	✓	✓	✓	✓
Recency of Use, Past Year But Not Past Month Use	✓	✓	✓	✓
Recency of Use, Past 3 Years But Not Past Year Use	✓			
12-Month Frequency of Use		✓	✓	✓
30-Day Frequency of Use	✓	✓	✓	
30-Day Binge Drinking Frequency		✓		

[1] "Tobacco products" description contains cigarettes, cigars, and smokeless tobacco (chewing tobacco and snuff). The imputation of pipes was completed in the univariate step because only two recency categories (past month and not past month) and no frequency-of-use variables were available for pipes.

6.5.5.3 Final Recency-of-Use and Frequency-of-Use Variables

As with all other imputation-revised variables, the final imputation-revised recency-of-use and frequency-of-use variables were identified with the prefix IR, followed by a 5-letter identifier, where a 3-letter code identified the drug[95] and the final 2 letters identified the measure (RC = recency; FY = frequency of use in past 12 months; FM = frequency of use in past 30

[95] The exception to this rule occurred with marijuana, which for historical reasons contained only a two-letter code (MJ). Marijuana variables therefore ended with a four-letter identifier, rather than five.

days). Each IR-variable was accompanied by two imputation indicators, one with an II prefix and the other with an II2 prefix. The levels for the II-indicator were the standard levels used for all imputation-revised variables: 1 = questionnaire data; 2 = logically assigned; 3 = statistically imputed; and 9 = legitimate skip (where applicable). The II2-indicators contained more details, including information from the lifetime usage imputations indicating whether lifetime usage was imputed. The imputation indicator levels are provided in Table 6.8.

Table 6.8 Detailed Imputation Indicators for Recency and Frequency of Use

Level	Measure	
	Recency of Use	**Frequency of Use**
1	Questionnaire data	Questionnaire data
2	Logically assigned	Logically assigned[1]
3	Lifetime usage imputed	Lifetime usage imputed
4	Edited recency = 9 (lifetime user)	Lifetime usage not imputed
5	Edited recency = 8 (past year user)	N/A
6	Edited recency = 19 (lifetime not past month user)	N/A
7	Edited recency = 14 (lifetime not past year user)	N/A
9	N/A	Legitimate skip

[1] The logically assigned cases for 12-month frequency of use were not all included in Level 2. Some were included in Level 1. This occurred if the 12-month frequency of use was trimmed due to (1) 30-day frequency; (2) estimated 30-day frequency; or (3) month and year of first use.

6.6 Age at First Use and Related Variables

Unlike the recency and 12-month frequency-of-use variables, age at first drug use was not statistically imputed in the surveys prior to the 1999 survey. Instead, missing values were excluded from subsequent analyses. However, as with the 30-day frequency, missing age-at-first-use values have been imputed 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. It is important to note that in addition to age at first use for cigarettes, those respondents classified as lifetime daily cigarette users also were asked their age at first daily cigarette use.

6.6.1 Age at First Use

The age-at-first-drug-use imputations followed the same general procedures as the imputation of other drug use measures. A linear regression model utilizing SUDAAN® software was fitted using a logit transformation of the respondent's age at first drug use as a proportion of their current age as the response variable. UPMNs were formed using the predicted mean from the regression model. Each item nonrespondent's neighborhood was restricted by logical constraints and likeness constraints. From these neighborhoods, a final imputation-revised age at first use was created. In addition, a randomly assigned date (i.e., year, month, and day) of first use was constructed that remained consistent with the imputed age at first drug use and other drug use measures.

6.6.1.1 Hierarchy of Drugs

The first step in the imputation of age at first use was to determine the order in which drugs would be modeled. As with the other drug use measures, it was expected that age at first use of other drugs would be strong predictors of age at first use of each drug of interest. Therefore, a hierarchy was chosen in order to get the greatest benefit from using the previously imputed age-at-first-use values as predictors for the drug of interest. The hierarchy for age at first use was identical to the lifetime and recency/frequency-of-use hierarchy shown in Table 6.3.

6.6.1.2 Setup for Model Building and Hot-Deck Assignment

As with the imputation of other drug use measures, the file was broken into three age categories for the imputation of age at first use (12 to 17 years, 18 to 25 years, and 26 years or older), and all subsequent procedures were performed separately within each of these age groups. To impute missing age at first use for each drug, it was necessary to define the eligible population. Using the imputed recency of use, the files were reduced to lifetime users for each drug. If a valid response was provided for the age-at-first-use measure,[96] the person was deemed an item respondent. Before modeling, the respondent weights were adjusted, using a response propensity model, to match the entire population of lifetime users. (Weights were defined in the same way as with other drug use variables. See the discussion in Section 6.3.2 about how the weights were defined.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The following categorical covariates were included in the models: imputed recency of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives (where available, otherwise lifetime indicators were used); categorical age;[97] race/ethnicity; gender; census region; and an MSA indicator.[98]

6.6.1.3 Sequential Model Building

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, since 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

---

[96] Respondents who reported age at first use of 1 or 2 were not included in the model.

[97] The covariate "categorical age" was divided into five categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). For the 12-to-17 and 18-to-25 age groups, categorical age was not included as a covariate in the item response propensity models.

[98] These variables were included in every model unless convergence problems arose. If this occurred, the model was reduced.

$$Y_i = Age\ at\ First\ Use_i + Uniform(0,1)\ random\ number.$$[99]

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. Variables included in the regression equation were modified 12-month and 30-day frequencies for the drug in question; modified versions of the imputed age at first drug use for previously imputed drugs; imputed recency of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives (where available, otherwise lifetime indicators were used); centered age;[100] centered age squared; centered age cubed; gender; race/ethnicity; State rank (based on the recency variable, see Section 6.5.1 for details); first-order interactions of centered age, centered age squared, gender, and race/ethnicity; marital status; education level; employment status;[101] census region; and an MSA indicator.[102] The modified variables for 12-month frequency of use (where applicable), 30-day frequency of use (where applicable), and age at first use were defined as follows:

new12$_i$	=	0	if respondent did **not** use the i^{th} drug in the past 12 months
	=	12-month frequency	if respondent used the i^{th} drug in the past 12 months
new30$_i$	=	0	if respondent did **not** use the i^{th} drug in the past month
	=	30-day frequency	if respondent used the i^{th} drug in the past month
AFU$_i$	=	0	if respondent is **not** a lifetime drug user of the i^{th} drug
	=	age at first use	if respondent is a lifetime drug user of the i^{th} drug

[99] 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.

[100] The covariate age was centered within each age group in order to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[101] Marital status, education level, and employment status were included as covariates for the 18- to 25-year-old and 26 or older age groups only.

[102] These variables were included in every model unless small sample sizes precluded the use of such a large pool of covariates. If this occurred, the model was reduced.

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 users of this substance. A summary of the final models can be found in Appendix F.

6.6.1.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

From the final model, a predicted value (based on the Y variable) was computed for each user of the drug of interest, which was then back-transformed to produce a predicted age at first use. The imputation-revised age-at-first-use assignment was conducted using the UPMN imputation described in Appendix C, where the "predicted mean" was the predicted age at first use. Again, this procedure defined a "neighborhood" of respondents by requiring that the respondents' predicted age-at-first-use values to have been within a certain relative distance, delta, of the nonrespondent's value. The value of delta was set so that donors were required to have a predicted age at first use within 5 percent of that of the item nonrespondent. If no donors were available with predicted means within 5 percent of the recipient's predicted mean, the neighborhood was abandoned and the respondent with the closest predicted age at first use was chosen as the donor.

6.6.1.5 Assignment of Imputed Values

Subject to the constraints described in the next section, separate assignments of provisional values were performed within each of the three age groups. The age at first use of the randomly selected donor was then transferred to the recipient.

6.6.1.6 Constraints on Univariate Predictive Mean Neighborhoods

As with all other drug use measures, imputations were conducted separately within each age group: 12 to 17, 18 to 25, and 26 or older. This could be considered a likeness constraint based on age, which was never loosened. In fact, recipients and donors were required to have been of the same age, if possible. If a donor of the same age was not found, the constraint eventually reduced to a logical constraint, where the imputed age at first use was less than the recipient's age. A small delta also could have been considered a likeness constraint, which could have been loosened by enlarging or removing delta. Initially, the relative distance for determining age at first use imputation neighborhoods (delta) was set so that any potential donor's predicted age at first use was within 5 percent of the recipient's predicted age at first use. Another likeness constraint, in addition to the match on age, required an approximate match on recency of use. The match was approximate because recipients who were past year users could have had donors who had used at any time in the past year (no distinction was made between past month and past year but not past month use). Finally, an attempt was made to require donors and recipients to be from States with similar usage levels, where usage was defined in terms of the prevalence of past month usage of the drug in question.

These 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. The order of loosening constraints occurred as follows: (1) removed the State rank group; (2) abandoned the neighborhood and chose the donor with the closest predicted mean; (3) loosened the restriction requiring an approximate match on recency of use and instead required only that recipients who did not use in the past year had donors who also did not use in the past year (tobacco recipients who did not use in the past 3 years had donors who did not use in the past 3 years); (4) loosened the restriction that donors and recipients had to have been the same age and instead required that the donor's age had to have been greater than or equal to the recipient's age and the donor's age at first use had to have been less than or equal to the recipient's age at first use;[103] and (5) loosened the "same-age" restriction even further so that the donor's age at first use was required only to have been less than or equal to the recipient's age. A summary of the above constraints and the number of respondents with sufficient donors corresponding to each likeness constraint are listed for each drug in Appendix G.

For drugs with no multivariate assignment, there were several logical constraints. For those respondents with an age at first use that equaled the recipient's current age, donors were excluded under the following circumstances. First, if the recipient's 12-month frequency was greater than the number of days since his or her last birthday, donors whose age at first use was equal to the recipient's current age were excluded. For example, suppose an item nonrespondent's birthday was on March 1 and the interview date was June 30. Then the number of days between the interview date and the respondent's birthday (inclusive) is 122. If the respondent had a 12-month frequency of 140 (either reported or imputed), his or her age at first use could not have been his or her current age. Second, if the respondent's recency of use indicated that he or she did not use in the past month, but the number of days since his or her last birthday was fewer than 30, the recipient's age at first use could not have been equal to his or her current age. And third, if the respondent was not a past month user, but the difference between his or her 12-month frequency and the days since his or her last birthday was fewer than 30, the recipient's age at first use could not have been equal to his or her current age. Consider again the example where the respondent's birthday was on March 1, the interview was on June 30, and the number of days between the interview date and the respondent's birthday (inclusive) is 122. If the respondent's recency of use indicated past year but not past month use, but his or her 12-month frequency was 111, some of those 111 days had to have occurred before his or her birthday, and the respondent's age at first use could not have equaled his or her current age. In addition, respondents with age-at-first-use values of 1 or 2 were not eligible to have been donors. Finally, cigarettes had yet another logical constraint: if the recipient was a daily cigarette user and his or her age at first daily use was not missing, the donors were prevented from having an age at first use later than the preexisting age at first daily use.

6.6.1.7 Multivariate Assignments

For smokeless tobacco (chewing tobacco and snuff), cocaine (crack), hallucinogens (LSD, PCP, and Ecstasy), pain relievers (OxyContin), and stimulants (methamphetamines), 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. Drugs where multivariate assignments were necessary are discussed in the following sections.

[103] With the loosening of the recency constraint, it was necessary to include a requirement that if the recipient was not a past year user, the age at first use could not have equaled the current age.

6.6.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

For reasons discussed in Section 6.3.7.1, one model for smokeless tobacco was fitted rather than individual models for chewing tobacco and for snuff. The nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted age at first use. Missing age-at-first-use values for chewing tobacco and/or snuff were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and, if both chewing tobacco and snuff were missing, imputed values came from the same donor. As for the constraints on the neighborhoods, all the constraints listed in the previous section were applied to both snuff and chewing tobacco separately. The likeness constraints also were applied to both chewing tobacco and snuff separately, but when loosened, they were loosened for chewing tobacco and snuff simultaneously. It is important to note that, for both chewing tobacco and snuff, lifetime usage was considered known (employing the lifetime usage imputation) so that there was no question of use versus nonuse of chewing tobacco or snuff. If age at first use was missing for chewing tobacco or snuff in the original data, but the respondent was imputed to have been a nonuser of chewing tobacco or snuff in the lifetime imputation, the respondent's age at first chewing tobacco use or age at first snuff use would have been adjusted to reflect the situation. Age at first use for smokeless tobacco was obtained by taking the minimum age at first use from chewing tobacco and snuff.

6.6.1.7.2 Cocaine and Crack

Even though cocaine and crack were in distinct modules, an age-at-first-use model was fitted for only cocaine. The nearest neighbor hot-deck neighborhood was then based on the overall predicted age at first use for cocaine. Missing age-at-first-use values for cocaine and/or crack were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and, if both cocaine and crack were missing, the imputed values came from the same donor. As for the constraints on the neighborhoods, all the constraints listed in the previous Section 6.6.1.6 were applied to both cocaine and crack separately. For example, donors for cocaine were logically restricted so that, if the recipient's 12-month cocaine frequency was greater than the number of days since his or her last birthday, donors whose ages at first cocaine use were equal to the recipient's age were excluded. The same was true for crack. The likeness constraints also were applied to both cocaine and crack separately, but, when loosened, they were loosened for cocaine and crack simultaneously. It is important to note that, for both cocaine and crack, lifetime usage was considered known (employing the lifetime usage imputation) so that there was no question of use versus nonuse of cocaine or crack. If age at first use was missing for crack in the original data, but the respondent was imputed to have been a nonuser of crack in the lifetime imputation, the respondent's age at first crack use would have been adjusted to reflect the situation.

Because crack is a type of cocaine, additional logical constraints were required so that donated values would have been consistent with preexisting nonmissing values. Specifically, if the crack age at first use was missing, but cocaine age at first use was not, the donated crack age at first use could not have been earlier than the preexisting cocaine age at first use. Conversely, if the cocaine age at first use was missing and crack age at first use was not, the donated cocaine age at first use could not have been later than the preexisting crack age at first use. Finally, if crack age at first use was missing, but the respondent was a crack user, the donor had to have been a crack user.

6.6.1.7.3 Hallucinogens (LSD, PCP, Ecstasy, and Other Hallucinogens)

The hallucinogens module consisted of many subgate questions, and three substances—LSD, PCP, and Ecstasy—were child drugs. One model was fitted for hallucinogens' age at first use, from which a single neighborhood was created for LSD, PCP, Ecstasy, and hallucinogens as a whole. The nearest neighbor hot-deck neighborhood was then based on the overall hallucinogens' predicted age at first use. Missing ages at first use for any or all of LSD, PCP, Ecstasy, and hallucinogens as a whole were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and, if any of the four ages at first use were missing, the imputed values came from the same donor. As for the constraints on the neighborhoods, the constraints listed in the previous section were all applied to hallucinogens as a whole. Because no 12-month frequency was available for the child drugs, it was not possible to implement any constraints on these drugs involving the 12-month frequency.

Because of the parent/child relationship, additional logical constraints were required so that donated values were consistent with preexisting nonmissing values. For example, if the ages at first use for LSD and PCP were missing, but the ages at first use for overall hallucinogens and Ecstasy were not, the donated LSD and PCP ages at first use could not have been earlier than the preexisting overall hallucinogens age at first use (but the LSD and PCP ages at first use could have been earlier than the Ecstasy age at first use). Another example is if the age at first use for hallucinogens was missing and the LSD age at first use was not (and the respondent was a nonuser of both PCP and Ecstasy), then the donated overall hallucinogens age at first use could not have been later than the preexisting LSD age at first use. In addition, if any of the child ages at first use were missing, but the respondent was a user, the donor also had to have been a user. Finally, if the respondent used one or more of the child drugs, but used no "other" type of hallucinogen, then his or her overall hallucinogens age at first use was imputed (or assigned) to have been equal to the minimum of the child ages of first use.

All of the constraints applied specifically to the child drugs were logical constraints. It is important to note that, for both the parent and child drugs, lifetime usage was considered known (employing the lifetime usage imputation) so that there was no question of use versus nonuse. If an age at first use was missing for one or more of the child drugs in the original data, but the respondent was imputed to have been a nonuser of any of these drugs in the lifetime imputation, then the respondent's age at first use of would have been adjusted to reflect the situation.

6.6.1.7.4 Pain Relievers (OxyContin and Other Pain Relievers)

For pain relievers, OxyContin was a child drug. One model was fitted for age at first use of pain relievers, from which a single neighborhood was created for both OxyContin and overall pain relievers. The nearest neighbor hot-deck neighborhood was then based on the overall pain relievers' predicted age at first use. Missing ages at first use for OxyContin and/or overall pain relievers were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and, if both OxyContin and overall pain relievers were missing, the imputed values came from the same donor. As for the constraints on the neighborhoods, the constraints listed in the previous section were all applied to overall pain relievers.

As for hallucinogens, additional logical constraints were required to account for the parent/child relationship. Specifically, if the age at first use for OxyContin was missing, but overall age at first use of pain relievers was not, the donated age at first use of OxyContin could not have been earlier than the preexisting age at first use of pain relievers. Conversely, if the age at first use of pain relievers was missing and the age at first use of OxyContin was not, the donated age at first use of pain relievers could not have been later than the preexisting age at first use of OxyContin. In addition, if the age at first use of OxyContin was missing, but the respondent was an OxyContin user, the donor had to have been an OxyContin user. Finally, if the respondent used OxyContin, but used no "other" type of pain reliever, then the overall pain reliever age at first use was imputed (or assigned) to have been the same value as the OxyContin age at first use. All of the constraints applied specifically to OxyContin were logical constraints. It is important to note that, for both pain relievers and OxyContin, lifetime usage was considered known (employing the lifetime usage imputation) so that there was no question of use versus nonuse of OxyContin. If age at first use was missing for OxyContin in the original data, but the respondent was imputed to have been a nonuser of OxyContin in the lifetime imputation, then the respondent's age at first use of OxyContin would have been adjusted to reflect the situation.

6.6.1.7.5 Stimulants (Methamphetamines and Other Stimulants)

The handling of the age–at-first-use variables for methamphetamines and overall stimulants was very similar to the procedures for OxyContin and overall pain relievers, as described in the previous section.

6.6.1.8 Year of First Use, Month of First Use, and Day of First Use Assignments

After the age-at-first-use imputations, all lifetime users of a given drug had nonmissing age-at-first-use values. Using this age at first use (AFU), users were assigned year/month/day of first use values. 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:

1. For recent initiates: Missing day of first use only
2. For recent initiates: Missing month/day of first use
3. For recent initiates: Missing year/month/day of first use
4. For less recent initiates: Missing year/month/day of first use

For each missingness pattern, bounds on both the earliest possible date of first use and the latest possible date of first use were determined. The final earliest possible date of first use was equal to the maximum of its bounds, and the final latest possible date of first use was the minimum of its bounds. Once the earliest and latest possible dates of first use were determined, a day was randomly selected from this interval. The imputation-revised month/day/year values were then extracted from this date of first use.

6.6.1.8.1 Missingness Pattern 1

In this case, the respondent 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 12-month 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 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* = *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 – 12-month frequency + 1), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)] *if recency* = 1,
>>
>> minimum [(Interview date – 29 – 12-month frequency), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)] *if recency* = 2, or
>>
>> minimum [(Interview date – 1 year), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)] *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. Since 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 had 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 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) Recency = 1.
4) 12-month frequency = 30-day frequency (if applicable), or 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 will not cause an inconsistency. The second condition implies that the drug was not used by these respondents more than 1 year ago. The third and fourth conditions imply that the drug was not used by these respondents in the interval (1 year before the interview, 1 month before the interview). Therefore, these respondents did not use the drug more than 1 month ago. All their lifetime use must have occurred in the past month.

6.6.1.8.2 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."

6.6.1.8.3 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 12-month 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.

6.6.1.8.4 Missingness Pattern 4

The fourth type of 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 frequency (or frequencies) was immaterial to the final date of first use because the respondent could not have begun using in the past year:

Earliest possible date = AFUth birthday, and

Latest possible date =

1 day before the (AFU + 1)th birthday if recency < 4, or

minimum [(Interview date – 3 years), (1 day before the (AFU + 1)th birthday)] *if recency* = 4.

6.6.1.8.5 Exceptions to the Standard Assignment of the Date of First Use

Although most of the drugs followed the standard assignment of the date of first use, a few exceptions occurred. These are listed below.

The tobacco products (cigarettes, cigars, chewing tobacco, and snuff) did not have a 12-month frequency. As a result, the 30-day frequency was used whenever possible. This affected only the *latest possible date*, which was defined as follows for these drugs:

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,

minimum [(Interview date – 1 year), (1 day before the (AFU + 1)th birthday)]
if recency = 3, and

minimum [(Interview date – 3 years), (1 day before the (AFU + 1)th birthday)]
if recency = 4.

Another variation occurred with the smokeless tobacco date of first use. In this case, 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, methamphetamines, 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 example, respondents with LSD recency = 3 could not have first used hallucinogens within the past year, regardless of the hallucinogens recency value. The bound effected by the child drug recency and frequency was calculated in exactly the same way as for the parent recency and frequency information (see Section 6.6.1.8.1).

For hallucinogens, pain relievers, and stimulants, an indicator of lifetime use of drugs other than the child drugs was created (see Table 6.2). 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, since there was more than one of them. The algorithm was as follows:

1) Assigned the date of first use for overall hallucinogens.

2) Assigned *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.

3) For respondents who were lifetime nonusers of other hallucinogens:

a. Determined which, if any, child drug could have had the same date of first use as hallucinogens. Specifically, determined whether the date of first use for hallucinogens was between *earliest possible date* and *latest possible date* for each child drug.

b. 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 with the hallucinogens date of first use was overwritten.

6.6.1.8.6 Final 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[104] 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).

6.6.2 Age at First Daily Cigarette Use Imputations

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 cigarette daily use were similar to age at first use, with two key exceptions.

The first 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 respondent had a "yes" value for the edited variable associated with the "ever-daily-used" question (CIGDLYMO),[105] he or she was considered a daily user. For more information about CIGDLYMO, see Kroutil et al. (2007). At this stage in the process, there should have been no missing responses to the 30-day frequency question. Daily users, based on 30-day frequency, should have been either known (based on a response in the survey) or imputed. However, missing responses for the ever-daily-used question also should have been imputed. The second exception involved the predicted means. Due to the high correlation between age at first use and

[104] 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 six. The code for cigarette daily use was CD2, which differed from the general cigarette code of CIG. Details about cigarette daily use are provided in Section 6.6.2.9.

[105] For the first time in the 2005 survey, the edited variable CIGDLYMO, instead of the raw questionnaire variable CG15, was used in the procedures for age at first daily cigarette use. This was done to facilitate the handling of a single respondent in the 2005 survey whose value for CIGDLYMO differed from the value for CG15. CIGDLYMO will be used instead of CG15 in future NSDUHs, as well.

age at first daily use, models for age at first use were used to define the imputation neighborhoods for age at first daily use.

Thus, the age-at-first-daily-use imputation involved two parts: The first part involved missing values in the ever-daily-used variable (CIGDLYMO). The second part involved all missing age-at-first-daily-use values for eligible daily users, including those that were imputed to have ever used daily.

6.6.2.1 Setup for Model Building—Ever-Daily-Used Variable (CIGDLYMO)

Because age at first daily use was asked of all persons who answered the ever-daily-used question with a "yes," it was necessary to ensure that this question had no missing values. As with all other drug use imputations, the file was broken into three age categories (12 to 17 years, 18 to 25 years, and 26 years or older), and all subsequent procedures were performed separately within these age groups. To impute for missing values in the ever-daily-used variable, it was necessary to define the eligible population—respondents who had an imputation-revised 30-day frequency[106] 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. Before modeling, the item respondent weights were adjusted to match the entire eligible population. This adjusted weight was computed using a response propensity model (see Appendix B for the more general GEM) and included the following categorical covariates: imputed recency of use for cigarettes; the lifetime indicators for smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; categorical age;[107] race/ethnicity; gender; census region; and an MSA indicator.

6.6.2.2 Model Building—Ever-Daily-Used Variable (CIGDLYMO)

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. Variables included in the initial regression equation were a revised 30-day cigarette frequency variable (in the same format as used in the age-at-first-use models; see Section 6.6.1.3); the imputation-revised cigarette age at first use; imputed recency of use for cigarettes; the lifetime indicators for smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; centered age; centered age squared; centered age cubed; gender; race/ethnicity; State rank (based on the recency variable); first-order interactions of centered age, centered age squared, gender, and race/ethnicity; census region; an MSA indicator; marital status; education level; and employment status.[108]

---

[106] 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).

[107] The covariate "categorical age" was divided into five categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). For the 12-to-17 and 18-to-25 age groups, categorical age was not included as a covariate in the item response propensity models.

[108] Marital status, education level, and employment status were included as covariates for the 18-to-25 and 26-or-older age groups only.

6.6.2.3 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Variable (CIGDLYMO)

From the final model, a predicted mean of the ever-daily-used variable was computed for each eligible respondent. The assignment of imputation-revised ever-daily-used values was conducted using UPMN imputation, as described in Appendix C, where the "predicted mean" was the predicted probability of daily use at some point in the respondent's lifetime, given that the respondent was a lifetime user, but not a current daily user. Again, the procedure defined a "neighborhood" of respondents (i.e., potential donors) by requiring that a respondent's predicted ever-daily-used probability to have been within a certain relative distance, delta, of the nonrespondent's predicted probability. Delta was set so that donors were required to have a predicted probability within 5 percent of that of the item nonrespondent.

6.6.2.4 Assignment of Imputed Values—Ever-Daily-Used Variable (CIGDLYMO)

Separate assignments were performed within each of the three age groups, subject to the constraints described in the next section. The ever-daily-used response of the randomly selected donor was then transferred to the recipient.

6.6.2.5 Constraints on Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Variable (CIGDLYMO)

As with all other drug use measures, neighborhoods for the ever-daily-used variable were restricted so that candidate donors and recipients were in the same age group (12 to 17 years, 18 to 25 years, or 26 years or older). Models were built separately within these three groups, so this likeness constraint was never loosened. The likeness constraints were nearly identical to those of age at first use (see Section 6.6.1.6). The only difference was in the definition of the predicted mean, the determination of which was described in Section 6.6.2.2.

6.6.2.6 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—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 UPMN imputation. The procedure defined a "neighborhood" of respondents by requiring that the respondent's predicted mean be within a certain relative distance, delta, of the nonrespondent's predicted mean.

6.6.2.7 Assignment of Imputed Values—Age at First Daily Cigarette Use

Separate assignments were performed within each of the three age groups, subject to the constraints described in the next section. The age at first daily use of the randomly selected donor was then transferred to the recipient.

6.6.2.8 Constraints on Univariate Predictive Mean Neighborhoods—Age at First Daily Cigarette Use

As with all other drug use measures, neighborhoods for age at first daily use were restricted so that candidate donors and recipients were in the same age group (12 to 17 years, 18 to 25 years, or 26 years or older). The likeness constraints were nearly identical to those used for age at first use (see Section 6.6.1.6). 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 both were mutually consistent.[109] A summary of the above constraints and the number of respondents who fitted into each one are listed for each drug in Appendix G.

All the logical constraints applied to age at first cigarette use also were applied to age at first daily cigarette use. In other words, simply replace the words "age at first use" with "age at first daily use" in Section 6.6.1.6. Besides those logical constraints, an additional logical constraint was applied specifically to age at first daily cigarette use. If the age at first use for a recipient with a missing age at first daily use was not missing, the donors were prevented from having an age at daily first use earlier than the preexisting age at first use.

6.6.2.9 Date of First Daily Cigarette Use Assignments

After the imputation-revised age at first daily cigarette use was created, all daily cigarette users had a valid age of first daily cigarette use. From this age, a year/month/day of first daily use was assigned. The date assignment procedure was identical to the procedure described in Section 6.6.1.8, using the same exceptions noted in Section 6.6.1.8.5 for tobacco products and child drugs.

[109] In the 2005 survey, the situation where no donors were available, even after loosening all constraints, never occurred. It has occurred in past NSDUHs, however, and the programming code still exists in case the situation occurs in future NSDUHs.

7. Nicotine Dependence

7.1 Introduction

The method used to measure dependence on nicotine in the 2005 National Survey on Drug Use and Health (NSDUH)[110] was first introduced in the 2001 survey and also was used in the 2002-2004 NSDUHs. The questions used in the 2005 survey were the same as those asked in other surveys since the 2001 NSDUH. As in the 2004 survey, only respondents who reported use of cigarettes in the past 30 days were asked these questions.

The method for determining nicotine dependence involved the calculation of a continuous scale, called the Nicotine Dependence Syndrome Scale (NDSS) (Shiffman, Hickcox, Gnys, Paty, & Kassel, 1995; Shiffman, Waters, & Hickcox, 2003). This scale was calculated from 17 NSDUH questionnaire items (see Table 7.1) that were asked of respondents who used cigarettes in the past 30 days. For a response to have been considered valid, an answer of either "1 = Not at all true," "2 = Somewhat true," "3 = Moderately true," "4 = Very true," or "5 = Extremely true" was required for each of the 17 questions. The scale was the mean value (appropriately adjusted where necessary) of the responses to the 17 questions, provided all 17 responses were nonmissing.

Of the eligible respondents who did not answer every one of the 17 questions, the majority were either missing a response from only one of the questions or did not answer any of the 17 questions. For the respondents missing only one of the 17 variables, imputation was used to fill in the values for the missing variable, using the information from the other 16 nonmissing variables through weighted least squares regression models. This resulted in 17 regression models, one for each variable. Weighted least squares regression was not entirely appropriate for these data, since both the response variable and the covariates were ordinal variables, and least squares methods generally require the data to be continuous. However, the scale was calculated as a mean from ordinal variables, and the imputed values were used as only one value out of 17 in the calculation of an arithmetic mean. Any bias that might have resulted from using an inappropriate type of model would have had a minimal effect on the resulting NDSS.

The imputations described in this chapter are almost unique in this report because they were not performed using the predictive mean neighborhood (PMN) technique described in Appendix C. Also, the NDSS mean value was calculated from edited versions of the 17 nicotine-dependence questionnaire variables. The majority of the editing procedures for these variables are described elsewhere (Kroutil, Handley, Suresh, Felts, & Bradshaw, 2007).

7.2 Edited Nicotine Dependence Variables

Table 7.1 shows the correspondence between the 17 questionnaire items used in the NDSS and the corresponding edited variables. Among eligible respondents (those who had used

---

[110] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

cigarettes in the past 30 days), the valid responses for the edited variables, as with the raw variables, were given as "1 = Not at all true," "2 = Somewhat true," "3 = Moderately true," "4 = Very true," or "5 = Extremely true" and were required. For most nicotine dependence variables, "dependence" was marked by the "Extremely true" response. However, for question variables DRCGE04, DRCGE12, DRCGE13, and DRCGE14, "dependence" was marked by "Not at all true."

Table 7.1 Mapping of Raw Nicotine Dependence Question Variables to Edited Variables

Question Variable	Question Text	Edited 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

7.3 Imputation-Revised Nicotine Dependence Variables

7.3.1 Setup for Model Building

In general, imputation models for variable types other than nicotine dependence in the 2005 survey were modeled sequentially so that variables that were modeled early in the sequence could be used as covariates in models for variables later in the sequence. This was done to avoid fitting separate models for each missingness pattern. In the case of nicotine dependence, however, no imputation was performed if more than one NDSS variable was missing. As a result, for each respondent where imputation could have been performed, all 16 nonmissing NDSS variables could have been used as covariates in the model for the 17th missing variable. Therefore, no sequential modeling was necessary. Item respondents therefore had to have complete data for all 17 of the NDSS questions used in the models, and logically they had to have used cigarettes in the past 30 days. Item nonrespondents were those who used cigarettes in the past 30 days and answered only 16 of the 17 NDSS questions with valid nonmissing responses. Respondents who had used cigarettes in the past 30 days and were therefore eligible to answer the NDSS questions but answered only 15 or fewer of those questions were left out of the modeling process. The missing values in the NDSS variables for these respondents remained missing in the imputation-revised variables. No response propensity adjustments were performed for the item respondent weights used in any of the models. However, the ratio-adjusted design-based weights were used in the imputation models. The variables included in the models are discussed in the next section.

7.3.2 Model Building

In the 2005 survey, one model was created for each NDSS variable. The response variable for each model was the edited variable that corresponded to the question text shown in Table 7.1. The covariates in each model were the remaining NDSS variables. For example, if CIGIRTBL was the response variable, then the covariates would be the remaining 16 NDSS variables: CIGCRAVE, CIGCRAGP, CIGINCTL, CIGAVOID, CIGPLANE, CIGRNOUT, CIGREGDY, CIGREGWK, CIGREGNM, CIGNMCHG, CIGSVLHR, CIGINFLU, CIGNOINF, CIGINCRS, CIGSATIS, and CIGLOTMR.

7.3.3 Computation of Predicted Means

If a respondent was missing only one of the 17 NDSS items, the predicted mean for this item was obtained using the coefficients corresponding to the other 16 nonmissing covariates from the appropriate weighted least squares regression. The covariates and the response variables were all ordinal, so it was possible for a predicted mean to have exceeded 5 or been less than 1. Section 7.2 describes the five valid responses.

7.3.4 Assignment of Imputed Values

For those respondents missing only one of the 17 NDSS items, the missing value was replaced by the predicted mean in the imputation-revised variable. No attempt was made to round the predicted mean, and no attempt was made to add a residual. The nicotine dependence imputation-revised variables were unique in that missing values remained as missing values if the respondent was eligible to answer the nicotine dependence questions but two or more NDSS items were missing. For the remainder of respondents, of course, the edited valid response was assigned.

7.4 Summary Information for Nicotine Dependence Variables

Imputations were necessary for the nicotine dependence variables to create an NDSS score for as many eligible people as possible. The imputation method was devised to be simple and easy to implement, given the complexities of handling this type of missing data. To avoid complicated models, imputations were limited to cases where the respondent answered 16 of the 17 questions. If an eligible respondent answered fewer than 16 questions, no imputations were performed. It was possible that the respondent was eligible to answer the questions about nicotine dependence because he or she was imputed to have been a past month cigarette user. Table 7.2 summarizes the eligibility of respondents to answer the nicotine dependence questions and reasons why respondents were eligible or not eligible. Furthermore, among respondents who were eligible, this exhibit gives details about the amount of nicotine dependence data that was missing. Also, this exhibit provides information on whether the respondent was imputed to have been a past month cigarette user. Consequently, the respondent would have been eligible to have nicotine dependence data, but would have had missing data for all the nicotine dependence variables.

Table 7.2 Summary of Response Patterns for NDSS Variables

Number of Missing NDSS Variables	Past Month Smoker	Frequency
NOT ELIGIBLE TO ANSWER NICOTINE DEPENDENCEQUESTIONS: 50,264		
17	No (not imputed)	50,245
17	No (imputed)	19
ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES NOT IMPUTED: 169		
17	Yes (not imputed)	13
17	Yes (imputed)	18
2-16	Yes (not imputed)	138
ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES IMPUTED: 223		
1	Yes (not imputed)	223
ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, NO MISSING VALUES FOR DEPENDENCE VARIABLES: 17,652		
0	Yes (not imputed)	17,652

7.5 Corrected Versions of Summary of Response Patterns for NDSS Variables for Survey Years 2001-2004

The summary of response patterns for NDSS variables presented in the versions of this report for each of survey years 2001-2004 had some incorrect values.[111] The correct values are presented below.

Note that the corrected counts for survey years 2001-2003 are in a separate table from the corrected counts for survey year 2004. This is because for the 2001-2003 NSDUHs, questions on recency of use on bidis[112] and clove cigarettes were used to determine eligibility for the nicotine dependence questions. For details, see Grau et al. (2005). These questions did not appear in the 2004 or 2005 NSDUH.

Table 7.3 Summary of Response Patterns for NDSS Variables: Corrected Counts for Survey Years 2001-2003

Number of Missing NDSS Variables	Past Month Smoker	Past Month User Bidis or Cloves	Frequency		
			2001	2002	2003
NOT ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS: 50,099 (2001), 48,833 (2002), 48,949 (2003)					
17	no (not imputed)	no	50,075	48,810	48,932
17	no (imputed)	no	24	23	17
ELIGIBILITY TO ANSWER NICOTINE DEPENDENCE QUESTIONS UNKNOWN: 155 (2001), 97 (2002), 90 (2003)					
17	no (not imputed)	not known	155	93	90
17	no (imputed)	not known	0	4	0
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES NOT IMPUTED: 238 (2001), 169 (2002), 182 (2003)					
17	yes (not imputed)	no or not known	45	12	16
17	yes (imputed)	no or not known	14	16	13
17	no (not imputed)	yes	5	0	2
17	yes (not imputed)	yes	1	0	1
2-16	yes (not imputed)	no or not known	155	127	145
2-16	no (imputed or not imputed)	yes	9	10	1
2-16	yes (imputed or not imputed)	yes	9	4	4

[111] In later releases of the 2004 Imputation Report, the table contained correct values.

[112] Bidis, as described in the computer-assisted interviewing (CAI) questionnaire, are small brown cigarettes from India consisting of tobacco wrapped in a leaf and tied with a thread.

Table 7.3 Summary of Response Patterns for NDSS Variables: Corrected Counts for Survey Years 2001-2003 (continued)

Number of Missing NDSS Variables	Past Month Smoker	Past Month User Bidis or Cloves	Frequency 2001	Frequency 2002	Frequency 2003
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES IMPUTED: 266 (2001), 224 (2002), 260 (2003)					
1	yes (not imputed)	no or not known	234	203	248
1	no (imputed or not imputed)	yes	10	4	4
1	yes (not imputed)	yes	22	17	8
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, NO MISSING VALUES FOR DEPENDENCE VARIABLES: 18,171 (2001), 18,803 (2002), 18,303 (2003)					
0	yes (not imputed)	no or not known	16,664	17,475	17,363
0	no (imputed or not imputed)	yes	323	254	167
0	yes (imputed or not imputed)	yes	1,184	1,074	773

Table 7.4 Summary of Response Patterns for NDSS Variables: Corrected Counts for Survey Year 2004

Number of Missing NDSS Variables	Past Month Smoker	Frequency
NOT ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS: 49,521		
17	no (not imputed)	49,504
17	no (imputed)	17
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES NOT IMPUTED: 150		
17	yes (not imputed)	12
17	yes (imputed)	9
2-16	yes (not imputed)	129
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES FOR DEPENDENCE VARIABLES IMPUTED: 246		
1	yes (not imputed)	246
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, NO MISSING VALUES FOR DEPENDENCE VARIABLES: 17,843		
0	yes (imputed or not imputed)	17,843

8. Household Composition (Roster)

8.1 Introduction

This chapter summarizes the techniques used to edit inconsistent values in the household roster and the techniques used to create and impute missing values in the roster-derived household composition variables for the 2005 National Survey on Drug Use and Health (NSDUH).[113] In addition, this chapter summarizes the procedures used to create edited proxy variables. The proxy variables summarize the selection and identification of a relative of the respondent who lived in the respondent's household (according to the household roster), who was 18 years of age or older, and who answered the health insurance coverage and income questions for the respondent. Imputations were accomplished using the predictive mean neighborhood (PMN) technique described in Appendix C. The editing procedures implemented on the household roster are described in the following sections. Additionally, the procedures used to create respondent-level detailed roster variables, the roster-derived household composition variables, and the roster-based proxy variables are summarized in the following sections.

8.2 Household Roster Edits

8.2.1 Description of Household Composition (Roster) Section of Questionnaire

The introductory question to the household roster portion of the questionnaire (QD54) was interviewer administered. This question asked the respondent for information regarding the number of people living in his or her household, where allowable entries ranged from 1 to 25. If either the interviewer indicated that the respondent lived alone or the question was unanswered, the household composition (roster) section was skipped. However, if the interviewer indicated a household size greater than 1, the interviewer was then prompted to ask the respondent questions about the age, gender, and relationship to the respondent of every member of the household, starting with the household's oldest member and including the respondent. If a pair of respondents was selected in a household, the interviewer indicated which member of a respondent's household roster corresponded to the other selected pair member. The roster entry for the respondent was referred to as the "self" entry. In effect, the respondent completed a grid with the number of rows corresponding to the value entered in QD54. Table 8.1 shows an example grid where the number of persons in the household is four. In this example, the roster of the wife/mother is shown and the indicator variable shows that the son was selected as the other pair member. The possible relationship codes and specific relationship details are listed in Table 8.2.

[113] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

Table 8.1 Household Composition (Roster) Grid Example, QD54 = 4

Person #	Relationship to Respondent	Age in Years	Other Member Selected[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 to only respondents who were part of a pair. The other member selected could not have been the self because respondents were not interviewed twice. The other member selected was the roster member who had a value of "1" for this variable.

Table 8.2 Household Composition (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	

8.2.2 Household Roster Consistency Checks

To reduce the amount of editing required during the data processing stage, consistency checks were included in the Blaise program code.[114] Two types of consistency checks were employed in the household roster section of the questionnaire. These checks (1) compared a roster entry corresponding to the respondent with previously entered questionnaire information or (2) compared a roster entry against other roster entries or the respondent's roster age for internal consistency.

8.2.2.1 Comparisons with Previously Entered Questionnaire Information

In the 2001 survey, a consistency check was added to the household roster section of the computer-assisted interviewing (CAI). This check was triggered if the interviewer reported a different gender for the respondent in the household roster than was previously recorded in the

[114] The Blaise program is the computer program within the computer-assisted interviewing (CAI) instrument that was used to direct the respondent and interviewer through the questionnaire.

interview (question QD01). The interviewer was required to change either the roster entry or the gender that had been entered at the beginning of the interview. In the 2002 survey, a new consistency check involving the respondent's age was added. Not only was it necessary for the respondent's gender in the roster to match the questionnaire gender, but also for the respondent's age in the roster to match the age that had been entered in the nonroster part of the questionnaire (the Blaise variable CURNTAGE). For the age check, the interviewer could have either changed the respondent's age entered in the roster or overridden the consistency check and provided an explanation as to why the roster age did not match CURNTAGE. Both of these consistency checks involved the respondent's own entry in the roster (the "self" entry). If the consistency check for age was overridden, the value for age corresponding to the self may not have matched the questionnaire-edited age. Explanations given by the interviewer for overriding this particular consistency check were carefully reviewed. In rare cases, the final value for age (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. Additional details about how roster age was used are described in Chapter 4. Strategies for the more common situation, where the original value for AGE was not set to the roster age, are discussed in Section 8.2.4.

8.2.2.2 Internal Consistency Checks

Since the 2002 survey, internal consistency checks have been implemented for the household roster. These checks were triggered if any of the following conditions occurred:

1. The interviewer reported that the respondent had more than one spouse or partner or reported a spouse and a 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 a live-in 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 live-in 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 greater than 24 years older or younger than the respondent.

12. The interviewer reported that a household member was a grandchild of the respondent and the respondent was 30 years old or younger.

13. The interviewer reported that a household member was a grandparent of the respondent and the respondent was 60 years old or older.

In the 2005 survey, a new consistency check was added and this check replaced checks #12 and #13 above and was triggered if

14. 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 the consistency check was triggered, the interviewer changed either an age code or a relationship code in the roster to a more appropriate value. As a result of new consistency checks introduced each year, fewer edits to the roster are implemented each survey year.. Nevertheless, any edit that was invoked because of an override to a consistency check was carefully scrutinized. The relevant household rosters, as well as the explanation given by the interviewer for the override, were 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 the edit was not applied. If the interviewer's explanation was not considered legitimate, then the edit was applied. More details about roster edits are provided in Section 8.2.5. Explanations given by the interviewers for the overrides and evaluations of their legitimacy are provided in Appendix J.

8.2.3 Preliminary Roster Edits

To facilitate processing of the roster variables, a "roster-level" file was created in which the number of records per respondent was given by the household size in question QD54. If the respondent quit the interview after the household size question or in the middle of the roster questions, "dummy" records were created that corresponded to the missing household members.

8.2.4 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. In theory, these values should have matched CURNTAGE and QD01, respectively. Since the check involving gender was not allowed to be overridden, the gender for self in the roster always matched QD01, which was equivalent to IRSEX (see Chapter 4). For the consistency check comparing the respondent's roster age against CURNTAGE, the age of self in the roster should be close to the questionnaire-edited age, AGE (see Chapter 4 for a description of the methodology used to create AGE), especially if the respondent age was set to the roster age. Moreover, the interviewer was required to confirm with the respondent that the respondent was in fact the identified "self." However it was possible to have problems matching AGE with the age of self in the roster. The interviewer was able to override the consistency check for age of self for one of two reasons: (1) the self was misidentified and another roster member was the true self, but the interviewer insisted on not changing the entries, or (2) the interviewer correctly identified the self, but insisted that the correct age for the respondent was different than CURNTAGE and other evidence did not support this insistence (AGE was not set to the roster age, as discussed in Section 8.2.2.1). 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 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.[115] If a misidentified self was diagnosed, it was assumed that the interviewer used the roster member identified as the self, rather than the respondent, as the point of reference. Using the example shown in Table 8.1, if the respondent's son was used as the reference point, the relationship for the respondent became "mother" instead of "self" and the "husband" became "father." Under these circumstances, the code for self was set to missing, and the respondent's roster entries did not include a self. The remainder of relationship codes in the roster also was set to missing. In some cases, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point.

8.2.4.1 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 because the original relationship code would no longer make sense, since the reference person had been changed.

8.2.4.2 Salvaging Relationship Codes with a Misidentified Self

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

[115] 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.

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 live-in partner, the roster member listed as "self" was actually a spouse or live-in 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.

8.2.5 Roster Edits for Other Household Members

Relationship codes were edited if the relationship of the roster member was impossible based on age and gender in relation to the self. Edits of 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. Interviewers' explanations for overrides to these consistency checks were carefully examined to assess the legitimacy of the override as explained in Section 8.2.2. Some edits were "automatic" in the programming code, which meant that the interviewer was assumed to have been wrong 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 the override of the consistency check was correct and therefore, the edit was applied only if the interviewer's explanation was suspicious. Interviewer's explanations for overrides to consistency checks and evaluations of their legitimacy are provided in Appendix J.

For all of the edits described below, the frequency of the application of each edit in the 2005 survey is listed. In some cases, this frequency is given for special cases within the description of the edit. The total number of applications in the 2005 survey is provided in parentheses after the description of each edit. The frequency in parentheses does not include cases where an override to a consistency check occurred and the explanation to the override was considered legitimate.

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

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. In these cases, even though the relationship code was incorrect, the reference person for the relationship code was still the respondent.

1. 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"), 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 8.2.2.2, 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 pair cases. If two pair members were selected in the household, this error could have been observed by examining the roster entries of the other pair member. 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 and the number of children on the screener roster, it was apparent how a correction should be made. In this instance, the offending age was replaced with the value given by the pair member whose roster age and screener age agreed. (2005 survey: was not applied)

2. If two members were selected in a household, the roster age for the other member selected was commonly not the same as the questionnaire-edited age (AGE, defined in Chapter 4) of the other pair member. In this case, the roster age for the other member selected was changed to this questionnaire-edited age value. (2005 survey: applied 3,157 times, the age differences were only 1 or 2 years, or replaced a missing value, in 2,941 cases)

3. If two members were selected in a household, the roster gender for the other member selected was often not the same as the imputation-revised gender (IRSEX, defined in Chapter 4) of the other pair member. In this case, the roster gender for the other member selected was changed to this imputation-revised gender value. (2005 survey: applied 44 times)

4. In previous survey years, the relationship code for grandchild (9) and grandparent (11) were commonly confused. With the introduction of consistency checks (consistency checks #2 and #3 in Section 8.2.2.2), this did not occur in the 2005 survey. The following edit that 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 grandchild, then the relationship code was changed to grandparent. (2005 survey: was not applied)

8.2.5.2 Edits to Relationship Codes: Changes to Missing Codes

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent and no internally consistent value could be used to replace it. These edits were performed before the edits listed in Section 8.2.5.1 were completed. For respondents that had changes to their rosters due to the edits described below, changes to age and gender due to the edits in Section 8.2.5.1 were checked to make sure that they did not impact the decision to implement the edits below. The relationship code in these instances, as listed below, was set to a bad data code.

1. More than one roster member aged 15 years or older was listed as respondent's unmarried partner or as being the respondent's spouse. This situation should have been covered by consistency check #1 in Section 8.2.2.2. Six overrides that were considered legitimate occurred in the 2005 survey. (2005 survey: was not applied)

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 partner code set to bad data. This situation should have been covered by consistency check #1 in Section 8.2.2.2, but three overrides that were considered legitimate occurred in the 2005 survey. (2005 survey: was not applied)

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 in Section 8.2.2.2. No overrides to this consistency check were observed in the 2005 survey. This edit would have been automatic for respondents younger than 15 years old. (2005 survey: applied once)

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 in Section 8.2.2.2. An override that was considered legitimate did occur in the 2005 survey, though not with a respondent younger than 15 years old. This edit would have been automatic for respondents younger than 15. (2005 survey: was not applied)

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 in Section 8.2.2.2. No overrides to this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

6. The roster member was the respondent's biological mother, but was more than 60 years older than the respondent. (2005 survey: was not applied)

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 in Section 8.2.2.2. All four overrides to this consistency check were considered legitimate in the 2005 survey. (2005 survey: was not applied)

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 difference in age between the biological sibling and the respondent exceeded 25 years, then a consistency check was triggered (consistency check #11 in Section 8.2.2.2). (2005 survey: applied once)

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. (2005 survey: was not applied)

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 in Section 8.2.2.2. One override to this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

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 in Section 8.2.2.2. No overrides to this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

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 in Section 8.2.2.2. The override for the respondent or roster member younger than 15 observed in the 2005 survey was not considered legitimate. In this case, the in-law was changed to bad data based on the interviewer's comment. The consistency check also was applied to 15- or 16-year-old respondents or roster members. Two overrides to this consistency check involving 15- and 16-year-olds occurred in the 2005 survey. (2005 survey: applied once)

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. (2005 survey: was not applied)

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 in Section 8.2.2.2. No overrides to this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

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 in Section 8.2.2.2. No overrides to this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

16. The roster member was the respondent's sibling and the previous roster member was a parent, but the roster member's age was within 4 years of the age of the parent. If the sibling was a half- or step-sibling, an additional requirement was that there was only one parent. (2005 survey: was not applied)

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 in Section 8.2.2.2. Similarly, if the roster member was a "grandparent" who was younger than the respondent, then this situation was covered by consistency check #2 in Section 8.2.2.2. If there was difference of 30 or more years this was covered by consistency check #14. This edit was applied two times in the 2005 survey, with both situations involving overrides to this consistency check. (2005 survey: was not applied)

18. If the respondent had two parents, but both parents were listed as biological mothers or biological fathers, the roster genders of both roster members were set to missing. (2005 survey: applied eight times)

8.2.5.3 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Illogical Child Code)

In Section 8.2.5.2, 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. Scenarios where the illogical code was "child" are listed below.

1. The interviewer might have put a roster member after the respondent's parent in the household roster. If the relationship code for that roster member was given as "child," the relationship code was illogical if the age made it impossible for the roster member to have been the respondent's child. (See #4 in Section 8.2.5.2.) In fact, if more than one "child" was listed after the respondent's parent, each would have been 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. (2005 survey: applied once)

2. In some cases, the interviewer's entry for a roster member listed as child might have simply been a typographical error (where the "3" should have been a "4"). 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 8.2.5.2). 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 a partner, and the roster member was not 12 or more years younger than the respondent, then the relationship code was changed to sibling. (2005 survey: applied five times)

3. Both sides in a selected pair were respondents 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, then the relationship code was changed to sibling. (2005 survey: was not applied)

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/aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2), then no uncles/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. (2005 survey: was not applied)

8.2.5.4 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Illogical Spouse Code)

The interviewer also could have used a wrong reference person with spouse codes. This error occurred most frequently when a selected child had a parent with a spouse (the other parent) or live-in partner ("unmarried partner"). Rather than identifying this individual as a "parent" or "other nonrelative," the interviewer identified the roster member as a spouse or live-in partner of the child even though they intended for the point of reference to have been the child's parent rather than the child. This manifestation of the invalid 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 (see Section 8.2.2.2), provided either the respondent or the roster member was 16 or younger. If any of the edits below were applied because of an override to one of these consistency checks, then it is noted in the affected edit.

1. Both sides in a selected pair identified a spouse/live-in partner, but 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/live-in partner, and the other pair member's spouse/live-in partner was incorrectly identified. 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/partner. If the misidentified code was "spouse," then the code was changed to

"parent." However, if the misidentified code was "live-in partner," then the roster member may or may not have been considered the parent of the respondent. In most cases where the misidentified live-in partner was the respondent's parent's live-in partner, the code was changed to parent. The exception occurred when (1) the live-in partner of this respondent's parent was the other respondent selected in a pair, and (2) the live-in 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 a live-in partner of the respondent's parent. (2005 survey: applied twice, once involving partners and once involving a spouse, all changed to parent)

2. As in the previous edit, both sides in a selected pair identified a spouse/live-in 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/live-in 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/live-in partner" was unusually large, then on each side the misidentified spouse/partner should have been considered a spouse/partner of the respondent's parent. If the misidentified codes were both "spouse," the codes were both changed to "parent." As stated above, if the misidentified codes were both "live-in 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 the previous edit (#1). The same special code as in the previous edit was used to identify a live-in partner of the respondent's parent. Hence, the incorrectly identified "spouse/live-in partner" code was changed for each respondent in the pair, either to "parent" or to the aforementioned special code. (2005 survey: was not applied)

3. In this edit, only one side in a selected pair identified a spouse (not live-in partner), but the spouse was identified even though either (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 says 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, 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, the pair was a sibling-sibling pair, and the spouse code was a typographical error (meant to have been a sibling, with a code "4" instead of "5"). (2005 survey: was not applied)

4. Only one side in a selected pair identified a live-in partner, but the live-in partner was identified even though either (1) the respondent was younger than 15 or (2) the live-in partner was younger than 15. If the respondent listed one parent, but the other pair member listed two parents, the pair was a sibling-sibling pair and the relationship code was in reference to the parent's live-in partner, then 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 live-in partner was less than 15 years, the pair was a sibling-sibling pair, then the live-in partner code was changed to sibling. No overrides of this consistency check occurred in the 2005 survey. (2005 survey: was not applied)

5. Both sides in a pair identified the same household member as spouse or live-in partner. If the previous roster member on one of the sides was a sibling, then the spouse/live-in partner should be considered the sibling's spouse/live-in partner. The spouse/live-in 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/live-in partner belonged. To maintain proper counts, the spouse/live-in partner code for the youngest pair member was changed. (2005 survey: was not applied)

6. A spouse or live-in partner was identified even though (1) the respondent had one parent in the household who was the roster member listed before the "spouse/live-in partner"; (2) either the respondent was younger than 17 years old or the respondent was between 17 and 20 years old and the "spouse/live-in partner" was older than the respondent's parent; and (3) the respondent was more than 15 years younger than the "spouse/live-in 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 live-in partner, the "partner" of the respondent was considered the live-in partner of the respondent's parent. The code was changed to "parent." For a household member with a spouse code who was 16 years of age or younger, this edit should have been covered by consistency check #4 in Section 8.2.2.2. A single override of this consistency check occurred in the 2005 survey; the override was not considered legitimate. (2005 survey: applied once)

7. In cases where the respondent was younger than 15 years old, he or she identified a spouse/live-in partner, and the above edits did not apply, then 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/live-in 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 in Section 8.2.2.2. No overrides to these consistency checks were observed in the 2005 NSDUH data that were not already handled by other edits in this section. (2005 survey: was not applied)

8.2.5.5 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Illogical Sibling Codes)

If the relationship code was identified as the respondent's sibling, but the age difference between the roster member 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" exceeded 50 years, then the sibling relationship codes were referencing the respondent's children's relationships to each other. The relationship codes were therefore changed to "child." Age differences greater than 25 years among biological siblings would have been covered by consistency check #11 in Section 8.2.2.2. All five overrides to this consistency check observed in the 2005 NSDUH data were considered legitimate. The other cases were checked individually, with particular scrutiny being placed on age differences between 20 and 25 years. (2005 survey: applied three times)

8.2.5.6 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Illogical Grandchild Codes)

If the relationship code was identified as 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 check would have been covered by consistency check #3. If the age difference between the "grandchild" and the respondent was younger than 30, then this check would have been covered by consistency check #14 in Section 8.2.2.2. (2005 survey: was not applied)

8.2.5.7 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Illogical In-Law Codes)

An invalid reference code also occurred with in-laws. 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 invalid 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/aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2) then no uncles/aunts lived in the household. If a pair was selected, no nonimmediate family codes were found in either pair member's roster. In either case, the relationship code was set to parent. Otherwise, no certainty was associated with the relationship code, so this code was set to missing. (2005 survey: was not applied)

8.3 Creation of 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. Each of these attributes had a multiple of 25 variables corresponding to the maximum of 25 members of a household. Separate variables were created 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 brought together into four sets of variables, one set for each attribute. The edits listed in Section 8.2 were incorporated into the values of the detailed roster variables, called ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only), PRNTYP1-PRNTYP25 (type of parent: biological, adoptive, etc.), SIBTYP1-SIBTYP25 (type of sibling: biological, adoptive, etc.), CHDTYP1-CHDTYP25 (type of child: biological, adoptive, etc.), and TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither).

8.4 Creation of Household Roster-Derived Variables

After replacing faulty 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 household roster-derived variables shown in Table 8.3. If any information in the roster was missing, the roster-derived variable was set to missing. However, if some of the roster records for a respondent's household had missing data, then roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could have been imputed. Details on the imputation of the household roster-derived variables are provided in Section 8.5. 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. This was because the interviews for each pair member could have occurred at different times, resulting in possible differences in the household composition.

The respondent's household size was assumed to equal the total number of rostered people in the household, TOTPEOP, as shown in Table 8.3. The value of TOTPEOP was expected to equal the value of QD54 in most cases. However, in some cases, the original self was misidentified and no other roster members were close to matching the respondent's age and gender. In these cases, an extra roster member was added to correspond to the respondent (the self) so that the value of TOTPEOP was 1 greater than the value of QD54. For other cases, the respondent did not enter a value for QD54, and thus TOTPEOP and all the roster-derived variables were missing. Finally, it was possible that duplicate entries were put into the household roster so that the value of TOTPEOP would have been 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 by 1 from the value of QD54.

The variables KID17 (number of children in the household younger than the age of 18) and HH65 (number of people 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 were missing, as well, regardless of whether some of the roster members were eligible to have been part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in Section 8.5.6. However, if the roster member was missing a relationship code, but not an age, then that roster member was still eligible to have been counted in these variables.

Table 8.3 Household Roster-Derived Variables

Variable Description	Variable Name
Total number of rostered people	TOTPEOP
Number of people in household aged 17 or younger	KID17
Number of people in household aged 65 or older	HH65
Indicator of whether the respondent had family members in household	FAMSKIP
Number of respondent's family members in household	FAMSIZE
Number of respondent's family members in household 0 to 17 years old	KIDFAMSZ
Number of respondent's children in household 0 to 2 years old	NRBABIES
Number of respondent's children in household 3 to 5 years old	NRPRESCH
Number of respondent's children in household 6 to 11 years old	NRYUNGCH
Number of respondent's children in household 12 to 17 years old	NRTEENS
Number of respondent's children in household younger than or equal to 17 years old	NRCH0_17
Number of respondent's children in household 18 to 20 years old	NROLDRCH
Number of respondent's children in household 21 or older	NROLDCH
Number of roommates/housemates in household	NROOMATE
Indicator of presence of mother in household (12- to 17-year-olds)[1]	IMOTHER
Indicator of presence of father in household (12- to 17-year-olds)[1]	IFATHER
Indicator of presence of foster child in household 12 to 14 years old[2]	FSTRCHLD

[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 a poverty variable. It was necessary because Federal poverty guidelines do not consider foster relationships as "family." Furthermore, the 12 to 14 age group was used because poverty is defined for respondents aged 15 or older. Foster children younger than age 15 are considered unrelated individuals for whom poverty cannot be measured. For details, see http://www.census.gov/hhes/poverty/povdef.html.

The variable FAMSKIP was an indicator of whether the respondent's household contained 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 FAMSIZE (number of respondent's family members in the household) and KIDFAMSZ (number of respondent's family members in the household younger than the age of 18) were simple counts based on the relationships of the individuals to the respondent and the ages in the respondent's household roster. These variables were created to determine appropriate measures of poverty levels, using Federal poverty definitions. (See http://www.census.gov/hhes/poverty/povdef.html for details on the poverty definitions.) The definition of "family" for FAMSIZE and KIDFAMSZ was a little different than that used for other roster variables. To maintain consistency with Federal poverty definitions, foster relationships were not considered family relationships. If some of the roster members had missing ages or missing relationship codes, the values of FAMSIZE and KIDFAMSZ were set to missing even though some of the roster members might have been eligible to have been part of the count. In these instances, the imputed values for FAMSIZE and KIDFAMSZ were restricted based on the nonmissing information available in the roster as explained in Section 8.5.6.

Eleven other roster-derived variables were created that used both the age and relationship codes of the roster members. All of the roster-derived variables and their definitions are summarized in Table 8.3. 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.

8.5 Imputation of Household Roster-Derived Variables

Although 17 roster-derived variables were created from the edited roster, missing values were imputed for only 6 of these variables: TOTPEOP, KID17, HH65, FAMSKIP, FAMSIZE, and KIDFAMSZ. The missing values in these variables were imputed using the univariate predictive mean neighborhood (UPMN) technique, as described in Appendix C.

8.5.1 Hierarchy of Household Roster-Derived Variables

After editing the roster variables, the next step in the imputation of household roster-derived variables was to determine the order in which the variables should be modeled. Each roster-derived variable was expected to have a high association with the other five roster-derived variables. Hence, it was important to perform the imputations sequentially so that variables early in the series were used as covariates for subsequent variables, if needed. The order in which the roster variables were imputed is shown in Table 8.4.

Table 8.4 Household Roster-Derived Variables (in Order of Imputation)

Roster Variable	Edited Variable	Imputed Variable
Total number of rostered people	TOTPEOP	IRHHSIZE
Total number of children younger than age 18	KID17	IRKID17
Total number of people aged 65 or older	HH65	IRHH65
Indicator of whether the respondent has family members in household	FAMSKIP[1]	IRFAMSKP
Total number of respondent's family members in household	FAMSIZE	IRFAMSZE
Total number of respondent's family members in household younger than age 18	KIDFAMSZ	IRKIDFAM

[1] FAMSKIP was set to 0 if the roster had relationship codes of 2, 3, 4, 5, 6, 8, 9, 10, 11, and 13 as described in Table 8.2. FAMSKIP was set to 1 if no relationship codes were missing and the roster had codes of 1, 7, 12, and/or 14 as described in Table 8.2.

8.5.2 Setup for Model Building

Once the hierarchy of the roster-derived variables was established, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all roster-derived variables were conducted separately within the four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. Response propensity adjustments were then computed for each age group to make the item respondent weights representative of the entire sample. (Because the modeling of the final weight adjustments was not completed at the time of the roster imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.[116]) Item respondents were not defined across all roster categories. Hence, this adjustment was computed separately for each age group and for each variable. The covariates in the response propensity models were the same covariates as those used in the main model considered in the next section. The item response propensity model is a special case of the generalized exponential model (GEM).[117] Greater details of the GEM software are presented in Appendix B.

8.5.3 Sequential Model Building

The variables TOTPEOP, KID17, HH65, FAMSIZE, and KIDFAMSZ were assumed to have a Poisson distribution, and the parameters for the models were estimated using the LOGLINK procedure in SUDAAN® software.[118] The binary variable FAMSKIP was modeled using weighted logistic regression. The covariates in each model were continuous centered age,[119] continuous centered age squared, gender, race/ethnicity, imputation-revised roster-derived variables earlier in the sequence, region, population density, percentage Hispanic/Latino households in segment, percentage of owner-occupied households in segment, and (for TOTPEOP only) number of people in the household eligible for interviewing (from the preinterview screener). There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups, the additional covariates of marital status, education status, and employment status also were included.

8.5.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

From the final models, a predicted mean was computed for every respondent. The assignment of imputed values for the roster-derived variables was conducted using the UPMN technique described in Appendix C.

8.5.5 Assignment of Imputed Values

Separate assignments were performed within each of the four age groups. A univariate imputation was implemented for each of the roster-derived variables within each age group, using the predicted means from the appropriate models. Assignments were made within preset

---

[116] In subsequent text, the use of the word "weights" will refer to the ratio-adjusted design weights.

[117] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

[118] SAS®-callable SUDAAN® was used to fit the binary logistic regression models. Details about the LOGLINK procedure are discussed and additional references are provided in the *SUDAAN® Language Manual, Release 9.0* (RTI, 2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of RTI International.

[119] The covariate age was centered within each age group to reduce the effects of multicollinearity, particularly with the squared age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

bounds, as discussed in the next section. If no imputed values were available within the preset bounds, a random imputation was performed within those bounds.

8.5.6 Constraints on Univariate Predictive Mean Neighborhoods

A univariate imputation was implemented on each variable within each age group after predicted means from the models had been determined. In a general UPMN imputation, the neighborhood is restricted by two types of constraints: (1) logical constraints (which cannot be loosened) to make imputed values consistent with a nonrespondent's preexisting nonmissing values of other variables, and (2) likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible.

The logical constraints on the neighborhoods were sequentially based on the information already available in the roster and on roster-derived variables already imputed. The assignment of imputed values for KID17 was restricted within a lower and upper bound based on the value of IRHHSIZE and the nonmissing ages in the roster. For example, if a household roster had four members, with two members aged 18 or older, and one member was missing age and another member was 18 years or younger then KID17 would be missing. Thus, at least one child younger than age 18 would be in the household and two adults would be in the household. Hence, the assignment of KID17 in this example would be restricted between the values of 1 and 2. Likewise, HH65 was restricted within bounds in the same manner, using the variables IRHHSIZE and IRKID17 and the nonmissing ages in the roster. FAMSIZE was also restricted within bounds based on IRHHSIZE and the nonmissing ages in the roster. KIDFAMSZ was restricted within bounds using the variables IRHHSIZE, IRFAMSZE, and IRKID17 and the nonmissing ages in the roster.

Likeness constraints also were applied to the imputation of missing values in TOTPEOP, KID17, HH65, FAMSKIP, FAMSIZE, and KIDFAMSZ. The delta constraint[120] could have been considered a likeness constraint and been loosened by enlarging delta or abandoning the neighborhood altogether and taking the donor with the closest predicted mean. For TOTPEOP, delta was the only likeness constraint. If possible, donors and recipients for KID17 and HH65 were required to have the same household size (IRHHSIZE, the imputation-revised version of the household size variable). Also, FAMSKIP donors and recipients were required to have the same values for IRKID17 (the imputation-revised version of KID17) and marital status. For FAMSIZE and KIDFAMSZ, donors and recipients were required to have the same values for IRHHSIZE and IRKID17. KIDFAMSZ donors and recipients also were required to have the same respondent family member size (IRFAMSZE, the imputation-revised version of FAMSIZE). For KID17 and HH65, the household size likeness constraint was loosened after abandoning the neighborhood. For FAMSKIP, the marital status likeness constraint was never loosened even after abandoning the neighborhood. For FAMSIZE and KIDFAMSZ, the likeness constraint for household size and KID17 were loosened before abandoning the neighborhood. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in Appendix G.

[120] Appendix A provides a description of the delta constraint.

8.6 Proxy Variables

8.6.1 Introduction

The proxy portion of the questionnaire allowed the interviewer to determine whether there was another person in the household that was better suited than the respondent to answer the questions about health insurance coverage and income. As in previous survey years, for respondents in households with two or more members, respondents were asked to provide a roster of all people living in the household (including the respondent) and the relationship of the respondent to the other household members. If the household contained at least one adult related to the respondent, the respondent was asked questions to determine whether this other person (or one of these other people) might be a more suitable proxy. The questions concerned with proxy information in the 2005 survey were the same as those asked in the 2003-2004 surveys, but were slightly different from those asked in the 1999-2002 surveys. For all surveys since the 1999 NSDUH, whether or not a proxy could be selected was based on whether family members aged 18 or older were in the household roster. However, in the 2003-2005 surveys, the respondent was asked to choose a suitable proxy from a list of eligible family members based on their household roster. In the surveys prior to the 2003 NSDUH, the respondents were allowed simply to supply the relationship of their proxy regardless of their answers in the household roster.

8.6.2 Editing of Proxy Variables

All survey respondents were allowed to choose someone to be their proxy as long as the following conditions were met:

a. There was more than one person in the household.

b. The eligible person was a relative (not a boarder, roommate, or some other nonrelative).

c. The eligible person was aged 18 or older.

Table 8.5 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.

Table 8.5 Mapping of Raw Proxy Information Variables to Edited 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

8.6.2.1 Edited Indicator of Potential Proxies in Household (EDFAM18)

As described in Section 8.4, 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 ambiguous due to a missing household size or missing values in the roster, FAMSKIP could not be determined. As described in Section 8.5, missing values in FAMSKIP were imputed in the variable IRFAMSKP. A similar variable was created to identify households where the respondent's household roster included other family members 18 years of age 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 and "0" indicated otherwise.

8.6.2.2 Editing of 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 by the computer. 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). The two situations were (1) the respondent had not identified any adult family members in the household, but had non-family-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 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).

8.6.2.3 Editing of 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).

In addition to the above edits, more detailed rules were used to assign values to PRXRELAT, the edited variable corresponding to QP02. The value of QP02, which identified the proxy for the respondent, was chosen directly from the respondent's household roster. To assign a code for QP02, a subset of the respondent's roster (called a proxy roster) was created that included only adult family members. In the cases where there were a large number in the proxy roster, only the first nine adult family members listed in this roster were allowed for selection. Once the proxy roster was established, the number selected in QP02 was matched to the corresponding person in the proxy roster. The definitions of the levels of PRXRELAT are shown in Table 8.6.

Table 8.6 Assignment of Values for PRXRELAT, Based on Proxy Member Relationship

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 live-in partner	Live-in-partner	Male
10 = Female live-in partner	Live-in 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

8.6.2.4 Missing Values in EDFAM18 and IRFAM18

As in previous years, missing values in EDFAM18 were replaced by "imputed values" in the imputation-revised variable called IRFAM18. In fact, the values of IRFAM18 were derived directly from IRFAMSKP. If the missing value in FAMSKIP was imputed to a value of 1 in IRFAMSKP, this value would be copied to IRFAM18. The same was true for an imputed value of 0 even though it was possible that the respondent had family members in the household and none were adults. However, the variable IRFAM18 was technically not used, because missing values in EDFAM18 implied missing values ("no answer" codes of 98) for all of the proxy variables. The imputation indicator for IRFAM18 (IIFAM18) was in fact an indicator of whether the value in IRFAM18 was derived from IRFAMSKP and not a true imputation indicator.

9. Income

9.1 Introduction

As with most of the imputation-revised variables discussed in the previous chapters of this report, imputations for the 2005 National Survey on Drug Use and Health (NSDUH)[121] were accomplished using the predictive mean neighborhood (PMN) technique, as described in Appendix C. The edits applied to the income variables are described in Kroutil, Handley, Suresh, Felts, and Bradshaw (2007).

The imputation of income was separated into two phases. The first phase was known as the "binary variable phase" and involved the imputation of all the binary income variables, as well as the number of months on welfare. This included the "yes-no" questions about the sources of income for the respondent and for the respondent's family living in the respondent's household, the number-of-months-on-welfare question (the only nonbinary variable in the binary variable phase), and a "yes-no" question regarding whether the respondent's income or the respondent's family income (in the household) was $20,000 or more (including income from the sources referenced in the previous questions). The second phase of the imputation of income was known as the "finer category phase" and consisted of imputing more specific income categories for the respondent and the respondent's family in the household.

9.2 Binary Variable Phase

9.2.1 Order of Modeling Income Variables

The first step in the imputation of income variables was to determine the order in which the variables would be modeled. A motivation for using a hierarchy in PMN is given in Appendix C for drug use variables. For a model predicting whether a respondent had a given source of income, other sources of income were useful covariates. Following a provisional imputation of missing income values in the binary variable phase, the indicators earlier in the sequence were used as covariates for income models later in the sequence. Any imputed values in the income variables were considered temporary at this stage. This was because the final imputation was not implemented for income indicators until the modeling was completed for all income variables in the binary variable phase. The order in which the income indicators were imputed is given in Table 9.1.

9.2.2 Setup for Model Building

Once the hierarchy of income variables in the binary variable phase was established, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all income indicators were conducted separately within the four age groups: 12 to 17, 18 to 25, 26 to 64, and respondents 65 years or older. For an individual to be considered an

[121] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

item respondent for income variables in the binary variable phase, he or she must have had complete data for all of the questions included in this phase. These questions consist of social security, supplemental security, welfare payments and services, investments, child support, wages, other sources of income, food stamps, months on welfare, and total family income (less than $20,000 versus $20,000 or more). Response propensity adjustments were then computed for each age group to make the item respondent weights representative of the entire sample. (As with health insurance, the final analysis weights were used as weights. See Chapter 10 for further discussion.) Because item respondents were defined across all the income variables in the binary variable phase, this adjustment was computed only once per age group and then used in the modeling of income indicators. The item response propensity model is a special case of the generalized exponential model (GEM), which is described in greater detail in Appendix B. The covariates in the response propensity model were the same as those included in the main model, which is discussed in the next section.

9.2.3 Sequential Model Building

Beginning with social security, the probability that a family received income from a given source was modeled for item respondents, within each age group, using the nonresponse-adjusted weights. For the models, the parameters were estimated using logistic regression.[122] The response variable for each model was the edited combination of the pair of questionnaire variables associated with each income topic in the binary variable phase, the names for which are given in Table 9.1. The covariates in each model were centered continuous age,[123] centered age-squared, gender, race, provisional income indicators imputed earlier in the sequence, region, population density, percentage Hispanic/Latino households in the segment,[124] percentage non-Hispanic/Latino black/African American households in the segment, percent of owner-occupied households, imputation-revised number of adults in household, imputation-revised number of children in household, imputation-revised number of adults aged 65 years or older in the household, and a three-level State rank variable. There were also predictors that consisted of one-way interactions of centered age with race, centered age with gender, race with gender, centered age squared with race, and centered age squared with gender. For the three older age groups, the additional covariates of marital status, education status, and employment status were used. For the State rank groups, definitions were determined in terms of the proportion of a given State's residents having an income greater than or equal to $20,000.

---

[122] In the 2005 NSDUH, the logistic regression models were run in SAS®-callable SUDAAN® rather than SAS®. Both SAS® and SUDAAN® yield the same predictive means given the same set of covariates, but because SUDAAN® acknowledges the survey design, it gives correct values for the standard errors associated with each parameter estimate. Details about the logistic regression model and additional references can be found in the *SUDAAN® Language Manual, Release 9.0* (RTI, 2004). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of Research Triangle Institute.

[123] The covariate age was centered within each age group to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

[124] Segments were the first-stage sample units in the multistage 2005 NSDUH sample. Each segment consisted of a set of U.S. Census Bureau blocks.

Table 9.1 Order of Imputation of Income Variables in 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 Investment Income	FAMINT
Family Child Support Payments	FAMCHD
Family Wages	FAMWAG
Family Other Income	FAMOTH
Family Food Stamps	FSTAMP
Family Months on Welfare	WELMOS
Total Family Income[1]	FINC1

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

The same covariates were used for both the months on welfare variable and the binary total family income variable. For the months on welfare variable, weighted least squares regression was used, where the dependent variable was a standard logit,[125] such that $Y = \text{logit}(p)$ and p = number of months on welfare divided by 12. The binary total family income variable was modeled using weighted logistic regression. For a complete summary of the income imputation models, see Appendix F.

9.2.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods

Following the modeling of each income variable in the binary variable phase, missing values were replaced by provisional imputed values. This was necessary so that these variables could be used as covariates in subsequent models. Although no provisional imputed values were used to build the models, it was necessary to calculate predictive means for all respondents, including item nonrespondents, using the parameter estimates from the models. This sometimes required the use of the provisional values for the covariates. The predicted probabilities from these models were used to assign provisional values using the univariate predictive mean neighborhood (UPMN) imputation method described in Appendix C.

9.2.5 Assignment of Provisional Imputed Values

Separate assignments of provisional values were performed within each of the four age groups (12 to 17, 18 to 25, 26 to 64, 65 or older) for all income variables. The final income imputations were multivariate across all the variables in the binary variable phase. These variables represented source of income, months on welfare, and total income. The multivariate imputation process is further described in Section 9.2.8.

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

9.2.6 Constraints on Univariate Predictive Mean Neighborhoods

After predictive mean values from the model had been determined, a univariate imputation was implemented on each variable within each age group. In general, the PMN is restricted by two types of constraints: (1) logical constraints (which cannot be loosened) to make imputed values consistent with a nonrespondent's preexisting nonmissing values of other variables, and (2) likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible. As a logical constraint in the binary income variable imputations, donors were required to have the same value for the family skip variable (IRFAMSKP) as the recipient. The neighborhoods for the binary income indicators were restricted so that candidate donors and recipients would have been within the same age group (12 to 17 years, 18 to 25 years, 26 to 64 years, 65 years or older). Models were built separately within these four groups, so this likeness constraint was never loosened. A small delta could also have been considered a likeness constraint, and it could have been loosened by enlarging delta or abandoning the neighborhood altogether and taking the donor with the closest predictive mean. More details about delta are described in Appendix C. This was the only likeness constraint that could have been loosened with the binary income provisional imputations.

9.2.7 Multivariate Assignments

The predictive means were calculated with edited family income variables (see Table 9.1) as the response variables. For each variable, neighborhoods were created using scalar-predictive means from the appropriate model. With respect to these scalar-predictive means, a univariate methodology was used to determine the neighborhood. In most cases, three edited variables were associated with each predictive mean, so that missing values for these three variables required assignment of imputed values. Hence, even when determining the provisional imputed values using the univariate procedure, the assignment of imputed values was multivariate for all binary phase variables with two exceptions: food stamps and months on welfare. The variables associated with each of the models are given in Table 9.2.

Table 9.2 Imputation-Revised Personal and Family Income Variables

Income Model	Variables
Social Security	IRPSOC, IROFMSOC, IRFAMSOC
Supplemental Security Income	IRPSSI, IROFMSSI, IRFAMSSI
Welfare Payments	IRPPMT, IROFMPMT, IRFAMPMT
Welfare Services	IRPSVC, IROFMSVC, IRFAMSVC
Investment Income	IRPINT, IROFMINT, IRFAMINT
Child Support Payments	IRPCHD, IROFMCHD, IRFAMCHD
Wages	IRPWAG, IROFMWAG, IRFAMWAG
Other Income	IRPOTH, IROFMOTH, IRFAMOTH
Food Stamps	IRFSTAMP
Welfare Months	IRWELMOS
Total Family Income	IRPINC1, IRFINC1, IRFAMIN1

9.2.8 Multivariate Imputation

Sections 9.2.1 through 9.2.7 summarize the specifics of separating the set of binary income variables (in the 2005 NSDUH) into item respondents and item nonrespondents. These sections also describe model building, computation of predictive means, and the assignment of imputed values for these measures using a univariate predictive mean. In most cases, however, these univariate assignments were only provisional. The final imputed values for these income measures were obtained using neighborhoods built on a vector of predictive means using the multivariate predictive mean neighborhood (MPMN) technique as described in Appendix C. Consistent with the univariate imputations, the multivariate assignments were done separately within four age groups: 12 to 17, 18 to 25, 26 to 64, and respondents 65 or older.

For these source-of-income variables, a single months-on-welfare variable, and the binary total income variables, the collective distance between their conditional predictive means for a given incomplete data respondent and the complete data respondents was determined using a Mahalanobis distance[126] within each age group. As with other applications of MPMN, the predictive mean vector used in the Mahalanobis distance calculation included only variables that were missing for a given item nonrespondent. For the recipient, only missing values among the variables were replaced by the donor's values. For example, if the respondent was missing only a response for the other-family welfare payments question, then only the donor's other-family welfare payments response was given to the recipient.

The predictive mean that results from the months-on-welfare model was a logit of the proportion of the year received. This logit was transformed back into a proportion, which was the predictive mean used to match donors to each recipient. This meant that the proportion could have been treated as a probability, which in turn could have been multiplied by the probability of receiving welfare in the past year. Hence, the matching predictive mean could have been made conditional on the receipt of welfare in the past year, if necessary. More details about how the months-on-welfare predictive mean was made conditional on receipt of welfare in the past year are presented in Appendix H.

Candidate donors were restricted according to logical constraints, which could not be loosened. As with the univariate provisional imputations, donors and recipients were required, as a logical constraint, to have had the same value for the family skip variable. In addition, if a respondent was missing the months-on-welfare question, but was not missing one of the feeders to this question, the donor and recipient were required to have the same values for the nonmissing feeder question variables. For months on welfare, the feeder questions were those involving welfare payments or welfare services. Missingness patterns and the logical constraints imposed for the binary income variables are presented in Appendix H.

A number of likeness constraints also were imposed on the multivariate neighborhood for the binary income variables. The donors were usually restricted to those who were the same age as the recipient, or if that constraint was too restrictive, an age within 5 years of the recipient. There was a high degree of association between respondents who received welfare payments,

---

[126] See Appendix C for a definition of Mahalanobis distance. A definition can also be found in Manly (1986).

welfare services, and food stamps. There was also a high degree of association between respondents earning an income from investments and respondents who had high incomes, both of which were negatively associated with welfare, welfare services, and food stamps. Hence, if a recipient required imputation for one or more of these six variables (i.e., welfare payments, welfare services, food stamps, binary income, investment 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. If one of the pair of income variables (personal and other-family-member source of income, or personal and family income) was missing, the donor and recipient were required to have the same value for the nonmissing variable.

Some other likeness constraints corresponded to covariates that were highly correlated with the response, but these constraints often were not included in SUDAAN® models. This was due to near-empty cells when the variables were cross-tabulated, causing instability in the estimates. In particular, this affected the following personal and/or other-family-member binary source-of-income variables: welfare payments, welfare services, child support, wages, and social security. The welfare and child support variables were strongly related to whether children were in the household. Because the variable representing the number of kids aged 18 years or younger in the household was included in the models, the following likeness constraint was added: both the donor and recipient had to either have kids aged 18 years or younger in the household, or not, provided the age group was 18 or older. This constraint was applied if one or more of the source-of-income variables (either personal or other family) for welfare payments, welfare services, or child support were missing. Likewise, new likeness constraints were added for the wages variable, which was highly correlated with employment status and, for respondents 65 or older, whether someone aged 65 years or younger was in the household. Specifically, if the personal wages response was missing among respondents aged 15 years or older, donor and recipient both had to be working or not working. Among respondents 65 years old or over, if personal wages or other-family-member wages variables were missing, the donor and recipient both had to either have someone aged 18 to 64 in the household, or not.

Finally, if the other-family-member social security value was missing, both donor and recipient had to either have someone aged 65 or older in the household, or not. If insufficient donors were present, the constraints were loosened in the following order: (1) abandoned the neighborhood and chose the donor with the closest predictive mean; (2) removed the requirement that donor and recipient needed to have been of the same age, but required them to have been within 5 years of each other; (3) removed the requirement that the donor and recipient be within 5 years of age of one another; (4) removed the constraint that incorporated the association between the welfare, food stamps, investment income, and total income questions; (5) removed the months-on-welfare constraints regarding personal and other-family-member welfare payments and services, and replaced it with a less strict requirement that the donor and recipient's family welfare payments and services must be a match; and then (6) removed the number-65-or-older constraint among respondents missing social security information and the number-under-18 constraint among respondents missing welfare payments, welfare services, and child support information. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in Appendix G.

9.2.9 Binary Income Recode: GOVTPROG

The 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, but not discussed in this document, GOVTPROG is described here because it was used as a covariate in subsequent health insurance models. (See Chapter 10 for details on the imputation of missing values in the health insurance variables.)

9.3 Finer Category Phase

9.3.1 Hierarchy of Income Variables

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 three variables were all considered simultaneously using a failure time model, which is described in greater detail in Section 9.3.3. Because only one model was fit, no hierarchy was required.

9.3.2 Setup for Model Building

As with the variables in the binary variable phase, the imputations were conducted separately within the four age groups: 12 to 17, 18 to 25, 26 to 64, and respondents 65 or older. For an individual to be considered an item respondent for income variables in the finer category phase, he or she must have had complete data for both questions in this phase. Response propensity adjustments were then computed for each age group to make the item respondent weights representative of the entire sample, and the appropriately adjusted weights were used in the models. (As with health insurance and the binary income variables, the final analysis weights were used as weights. See Chapter 10 for further discussion regarding health insurance.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The variables included in the model, which predicted the probability of item nonresponse, were the same as those included in the main model. Greater details are given in the next section.

9.3.3 Sequential Model Building

The finer categories of income were modeled using the LIFEREG procedure in SAS/STAT® software.[127] 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 assumed for the response variable representing income can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

[127] Details about the LIFEREG procedure are discussed in the *SAS/STAT User's Guide, Version 8* (SAS Institute, 1999).

where $\mathbf{y}$ is a vector of observed responses, $\mathbf{X}$ is the matrix of covariates, $\boldsymbol{\beta}$ is the parameter vector, and $\boldsymbol{\varepsilon}$ is a vector of error terms. Particularly, 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, such as the NSDUH income intervals. The contribution of an individual with covariates in the matrix $\mathbf{X}$ 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.[128]

LIFEREG allowed several choices for the functional form of the parametric model that corresponded to the error distribution discussed earlier, including the two-parameter log-logistic, lognormal, gamma, and Weibull, and also the three-parameter generalized gamma. Each of these models was fit to each of the four age-group-specific datasets. 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.

Many of the covariates considered in the model for the finer category phase included the same covariates used in the binary variable phase. These covariates included centered continuous age, centered age squared, gender, race, region, population density, percentage Hispanic/Latino population, percentage non-Hispanic/Latino black/African American population, percentage owner-occupied households, imputation-revised number of adults in household, imputation-revised number of children in household, imputation-revised number of adults aged 65 years or older in the household, and a three-level State rank variable. As in the binary variable phase, the State rank groups in the finer category group were defined in terms of the proportion of a given State's residents whose incomes were greater than or equal to $20,000. For both phases, there were also predictors that consisted of one-way interactions of centered age with race, centered age with gender, race with gender, centered age squared with race, and centered age squared with gender. For the three older age groups, the additional covariates of marital status, education status, and employment status were used for both the binary variable phase and the finer category phase. Also, all imputation-revised income indicators considered in the binary variable phase were used as covariates for the finer category phase.

9.3.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods

As described in the previous section, the failure time model contained 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.

---

[128] Details about the model specifications for LIFEREG models are given in SAS Institute (1999, pp. 1761-1796).

Specifically, $\mathbf{X\beta}$ was a translation of the estimated mean of log income. Mean values were computed for both item respondents and item nonrespondents using the parameters from the failure time model. Subsequently, these values were used to assign imputed values using the UPMN imputation method, described in general in Appendix C.

9.3.5 Assignment of Imputed Values

Separate assignments of imputed values were performed within each of the four age groups for all finer category income variables. Only missing values were replaced by imputed values using the same donor for both personal and family finer income variables. The multivariate imputation process is further described in Section 9.3.7.

9.3.6 Constraints on Univariate Predictive Mean Neighborhoods

Donors and recipients were required to have the same values for both the binary personal and family income variables and the indicator of whether other family members were in the household (IRFAMSKP). In addition, if either of the personal income or family income finer category responses were nonmissing, donors and recipients were required to have the same values for the nonmissing variable. Finally, donors were required to have predictive mean values "close to" (within the delta distance) the recipient's predictive mean value. If insufficient donors were available using these constraints, the constraint involving nonmissing personal or family income finer category responses was loosened to a logical constraint. This logical constraint required the recipient's nonmissing value to be consistent with the donor's value for the other variable. Finally, if no donors were available, the neighborhood was abandoned, and the donor with the closest predictive mean to the recipient was chosen, subject to the logical constraints. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in Appendix G.

9.3.7 Multivariate Assignments

The predictive means were calculated using the edited (finer category) family income variables (see Table 9.1) as the response variables. For each family income variable, neighborhoods were created using scalar-predictive means from the appropriate model. The methodology for determining the neighborhood was therefore univariate in terms of these scalar-predictive means. Three edited variables were associated with each predictive mean, so that the missing values for the three variables required assignment of imputed values. Hence, even when determining the provisional imputed values using the univariate procedure, the assignment of imputed values was multivariate for all but two of the variables. For the 2005 NSDUH, the imputation-revised variable for the personal income variable was called IRPINC2, the family income variable with legitimate skips was called IRFINC2, and the family income variable without legitimate skips was called IRFAMIN2.

9.3.8 Finer Category Income Recode: INCOME and INCOME5

The recoded variable INCOME classified the families of respondents into four income levels: less than $20,000; from $20,000 to $49,999; from $50,000 to $74,999; and greater than or equal to $75,000. Another recoded variable (INCOME5) was created to take advantage of an extra level of income. This variable had five levels: the first three were equivalent to INCOME, but the last level of INCOME was separated into two levels: $75,000 to $99,999; and greater than or equal to $100,000. Both INCOME and INCOME5 were recodes of the variable IRFAMIN2. A variety of recoded variables were created, but are not discussed in this document. However, as with GOVTPROG, the variable INCOME is discussed here because it was used as a covariate in subsequent health insurance models (see Chapter 10 for details on the imputation of missing values in the health insurance variables). INCOME5, which is currently used for special requests, is discussed here because of its similarity to the INCOME variable, and the fact that in future NSDUHs, it might be used in place of INCOME in those instances where INCOME is currently used.

10. Health Insurance

10.1 Introduction

Two methods were used to create the final imputation-revised health insurance variables. The first method, referred to as the "old method," followed the general strategy used in previous iterations of the National Survey on Drug Use and Health (NSDUH).[129] Specifically, this method was implemented to create two general imputation-revised health insurance variables. The first variable was simply an imputation-revised version of the edited private health insurance variable. For the second variable, a recoded overall health insurance variable was created by combining information from the edited health insurance variables, and then missing values for that recoded health insurance variable were imputed. Because the health insurance questions in the survey changed every year between the 1999 and 2001 surveys, different versions of the overall health insurance variable were created for each of these surveys. These two versions of the health insurance variable were created using the questions available in questionnaires from the 2002 survey onwards, including the 2005 survey. Thus, a total of three imputation-revised health insurance variables were created from the 2005 survey using the old method.

In the second method used to create the final health insurance variables, also known as the "constituent variables method," missing values in each of the constituent edited health insurance variables were individually imputed. This method was processed in two stages, where the four specific imputation-revised health insurance variables were created in the first stage, followed by the creation of the imputation-revised "any other" health insurance variable in the second stage. In this method, the overall health insurance variable was created by combining information from the five constituent imputation-revised health insurance variables. Regardless of how the final health insurance variables were derived, imputations were performed using the same methodology, the predictive mean neighborhood (PMN) technique, as described in Appendix C.

10.2 Edited Insurance Variables

Table 10.1 shows the edited counterparts for some of the health insurance questionnaire (raw) variables. In the 2005 survey, the edited variables had the same values as the questionnaire variables, except that missing values were replaced by standard NSDUH missing value codes.

10.2.1 Edited Insurance Variables (Old Method)

In the old method, three health insurance indicators were created from these six variables (see Table 10.1). Two of them, INSUR and INSUR3, indicated whether the respondent had any health insurance. The third, PINSUR, indicated whether the respondent had any private health insurance. INSUR3, which was consistent with the variable of the same name created in the 2001 survey, was coded as "yes" if any one of the six variables listed in Table 10.1 were coded as

[129] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

"yes," and "no" if all six variables were coded as "no." The other overall insurance indicator, INSUR, was created to maintain consistency with the 1999 survey. Because the questions associated with CHIPCOV (Children's Health Insurance Program) and HLTINNOS (covered by any kind of health insurance) did not exist in the 1999 questionnaire, these two variables were excluded from the determination of INSUR. The INSUR variable was coded as "yes" if any of the other four variables listed in Table 10.1 were coded as "yes" and "no" if all four variables were coded as "no."[130]

Table 10.1 Mapping of Raw Health Insurance Variables to Edited Counterparts

Question Variable[1]	Question Text[2]	Edited Counterpart[3]
QHI01 QHI01v	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)
QHI02, QHI02v	Is the respondent covered by Medicaid or Medical Assistance?	MEDICAID (1 = yes, 2 = no)
QHI02A	Is the respondent currently covered by a Children's Health Insurance Program operated by your State of residence?[4] (Asked only of respondents aged 12 to 19)	CHIPCOV (1 = yes, 2 = no)
QHI03	Is the respondent currently covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)
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])

[1] The "v" questions were asked to verify the answer given in the previous question for respondents who were younger than 65 and a Medicare recipient or older than 65 and a Medicaid recipient.
[2] The questions provided in this table are abbreviated versions of those given in the questionnaire.
[3] Missing values in these edited values were represented by standard missing value codes. CHIPCOV was replaced in the final analytic file by CAIDCHIP, a combination of MEDICAID and CHIPCOV. See Section 10.2.2 for details.
[4] The questionnaire did not ask the question exactly in this way. It identified the specific program, depending upon the State of residence entered by the respondent.
[5] A respondent was assigned a legitimate skip for HLTINNOS if they answered "yes" or gave no answer to at least one of the other health insurance questions.

To create the variable for private health insurance, PINSUR, only the edited variable PRVHLTIN (whether the respondent was covered by private health insurance at the time of the survey) was used. Missing data for the edited variable PRVHLTIN were coded using the standard NSDUH missing data codes for "don't know," refused, and blank, whereas missing data for PINSUR were all coded as "98," which was a code for missing data. Except for the codes used to handle missing data, PINSUR and PRVHLTIN were equivalent. The variable PINSUR was created to maintain consistency with pre-1999 surveys, in which other variables also contributed to the indicator of coverage by private health insurance. All respondents with private

[130] 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 created in the surveys since 2000.

health insurance were considered to have health insurance. Therefore, respondents with private health insurance were a subset of the respondents who had health insurance.

10.2.2 Edited Insurance Variables (Constituent Variables Method)

In the constituent variables method, the editing process combined the variables MEDICAID (whether the respondent was covered by Medicaid or Medical Assistance) and CHIPCOV (whether the respondent was currently covered by a Children's Health Insurance Program) to create the variable CAIDCHIP, which indicated whether someone was covered by Medicaid or one of the State children's health plans. This variable and all the other edited variables in Table 10.1, 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 10.1. 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/no imputed value for "any other health insurance" even though the respondent was never asked QHI11 and HLTINNOS = "99," which was a legitimate skip code. Thus the variable ANYOTHER 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) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, or PRVHLTIN = 1[131]
	= 2 if CAIDCHIP = 2, MEDICARE = 2, CHAMPUS = 2, and PRVHLTIN = 2
	= 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
	= SKHLCCOV if SKHLCCOV = 2 or a missing value code

10.3 Imputation-Revised Health Insurance Variables (Old Method)

The old method of creating the final imputation-revised health insurance variables amounted to imputing missing values in the recoded variables (INSUR and INSUR3), as described in the previous section and in PINSUR. This resulted in the creation of three imputation-revised variables, two for overall health insurance (IRINSUR and IRINSUR3), and one for private health insurance (IRPINSUR).

10.3.1 Order of Modeling Health Insurance Variables (Old Method)

A multivariate predictive mean neighborhood (MPMN) imputation method for private health insurance and overall health insurance was implemented. However, respondents who answered "yes" to the private health insurance question also were logically covered by overall

[131] SKHLCCOV was coded as a 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.

health insurance. Therefore, it was not possible to use INSUR or INSUR3 as covariates in the PINSUR model, or vice versa.

10.3.2 Setup for Model Building (Old Method)

After determining the modeling order of the health insurance variables, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all three health insurance variables were conducted separately within four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older.

The next step involved creating one model for PINSUR and another for INSUR3. A respondent was considered an item respondent for health insurance only if his or her status was known for both private health insurance and overall health insurance as defined by INSUR3. To meet this criterion, the respondent must have had a valid "yes" or "no" response in PRVHLTIN (the edited variable corresponding to QHI06 [whether the respondent was currently covered by private health insurance]). In addition, he or she either must have answered QHI01, QHI02, QHI02A, QHI03, or QHI11[132] (see Table 10.1 for descriptions of these variables) with a valid "no" response, or answered "yes" to at least one of the six questions (including QHI06). This ensured that the interview respondent's status with respect to both overall health insurance (INSUR3 definition) and private health insurance was completely known. For example, if the interview respondent did not answer QHI01, but answered "no" to the other five questions, his or her status with respect to overall health insurance depended on the missing response to QHI01. However, if the respondent answered "yes" to any of the other five questions, the value of INSUR3 was already known to be "yes."

Note that it was possible for a respondent to have been defined as an item nonrespondent for INSUR3, but as an item respondent for INSUR. This occurred if a respondent gave valid "no" answers to QHI01, QHI02, QHI03, and QHI06, but he or she did not answer QHI02A or QHI11 (and did not give a valid "yes" answer to either of these). On the other hand, because the variables making up INSUR constituted a subset of those corresponding to INSUR3, an item nonrespondent for INSUR was necessarily an item nonrespondent for INSUR3. Moreover, an item nonrespondent for PINSUR was necessarily an item nonrespondent for INSUR3. Because missing values in all three variables (PINSUR, INSUR, and INSUR3) were imputed, an item respondent was defined based on the response to INSUR3.

To ensure that the weights adequately represented the population, the weights for item nonrespondents (as defined by INSUR3) were reallocated to item respondents using item response propensity models within each age group for the pair INSUR3 and PINSUR. (Because the modeling of the final weight adjustments was not completed at the time of the health insurance imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.)[133] The item response

---

[132] References to QHI01 and QHI02 naturally imply that if the respondent was younger than 65 and answered "yes" to QHI01, then he or she also answered QHI01v. Moreover, if the respondent was 65 or older and answered "yes" to QHI02, then he or she also answered QHI02v.

[133] In subsequent text, the use of the word "weights" will refer to the ratio-adjusted design weights.

propensity model is a special case of the generalized exponential model (GEM),[134] which is described in greater detail in Appendix B. The variables included in the model predicting the probability of item nonresponse were the same as those included in the main model, which is discussed in the next section.

10.3.3 Sequential Model Building (Old Method)

The probability that the respondent had health insurance (as defined by INSUR3) and the probability that the respondent had private health insurance were both modeled for item respondents, within each age group, using the nonresponse adjusted weights. The private health insurance model was created only for respondents who were known to have overall health insurance so that the predicted probability modeled was P(PINSUR = 1 | INSUR3 = 1). For the models, the parameters were estimated using logistic regression.[135] Each response propensity model included the following pool of predictors: centered age,[136] race/ethnicity, centered age squared, centered age cubed, gender, population density, percentage of housing in segment that was owner-occupied, percentage of Hispanics/Latinos in the segment, percentage of non-Hispanic/Latino blacks/African Americans in the segment, and household size. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups (i.e., 18 to 25, 26 to 64, and 65 or older), the additional predictors of marital status, education level, and employment status also were considered in each model.

10.3.4 Computation of Predicted Means (Old Method)

Using the parameter estimates from models for overall and private health insurance, predicted probabilities of having insurance were computed for both item respondents and nonrespondents. In other multivariate imputations, a hierarchy was required, where provisional imputations were performed on variables earlier in the hierarchy to be used as covariates in variables further down the hierarchy. A final multivariate imputation was then performed on all the variables in the hierarchy. However, because neither variable could have been used as a covariate in the model for the other variable, no provisionally imputed values were required.

10.3.5 Multivariate Imputation of Health Insurance and Private Health Insurance (Old Method)

The final imputed values for overall health insurance (using both the INSUR and INSUR3 definitions) and private health insurance were obtained using neighborhoods built upon a vector of predicted means. The vector had two elements: *P*(overall health insurance, as defined

---

[134] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

[135] In the 2005 survey, the software used for most imputation modeling was SUDAAN®. However, the logistic model for the old method of imputing health insurance variables used SAS® to maintain consistency with the practice of previous survey years. SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

[136] The covariate age was centered within each age group to reduce the effects of multicollinearity, particularly with the squared and cubed age terms. For more information on "centering" and "multicollinearity," refer to Draper and Smith (1981).

by INSUR3) and *P*(private health insurance | overall health insurance, as defined by INSUR3). For both overall and private health insurance, the imputation method used was the MPMN procedure. More details regarding this imputation method are presented in Appendix C. Similar to the response propensity models, the multivariate assignments were done separately within the same four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older.

A respondent was eligible to have been a donor for a given item nonrespondent if he or she had complete data across PINSUR, INSUR, and INSUR3 and was within the same age group. Logical constraints were placed on individuals who were missing one or two of the three indicators. Respondents who were missing either of the overall indicators, but did not have private health insurance, required donors who also did not have private health insurance.[137] If a respondent was missing only INSUR3, then INSUR must have been "no" because a "yes" value for INSUR would have necessarily meant that INSUR3 would have been "yes" and therefore nonmissing. Hence, donors also must have had a "no" value for INSUR. By the same token, if a respondent was missing only INSUR or was missing both PINSUR and INSUR, but not INSUR3, then INSUR3 must have been "yes" because a "no" value for INSUR3 would have necessarily meant that INSUR would have been "no" and therefore nonmissing. In this case, donors must also have had a "yes" value for INSUR3. Finally, respondents who indicated that they had health insurance, but were missing the private health insurance indicator, required donors who had some health insurance.[138] As a likeness constraint, the set of potential donors was then further restricted to have been the same age as the recipient. If no eligible donors were available who had the same age as the recipient, donors were sought with ages within 5 years of the recipient. Finally, donors were required to have had all applicable elements of the multivariate predictive mean vector "close to" (i.e., within the delta distance) the recipient's elements of the predictive mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predictive mean vector. Likeness constraints were loosened in the order listed above. The patterns of missingness for overall and private health insurance, the logical constraints imposed on the set of donors, and the frequency of occurrence of each missingness pattern are provided in Appendix H. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in Appendix G.

The full predictive mean vector contained elements for overall health insurance (as defined by INSUR3) and private health insurance (conditional on a "yes" response to the overall health insurance (INSUR3) indicator). 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 a respondent was missing INSUR, but not INSUR3, the predicted mean that was derived using INSUR3 was used. The portions of the full predictive mean vector used to create the MPMN neighborhoods for each missingness pattern, with

[137] Technically, this was not a logical constraint because there was no restriction on whether the respondent did or did not have health insurance. However, because all respondents with private health insurance had health insurance and the recipient did not have private health insurance, the distribution would have been skewed in favor of a "yes" indicator if these respondents were allowed to be donors.

[138] Again, this technically was not a logical constraint. However, because all respondents who did not have health insurance also did not have private health insurance and the recipient had health insurance, the distribution would have been skewed in favor of a "no" indicator if these respondents were allowed to be donors.

accompanying adjustments, are provided in Appendix H. The Mahalanobis distance[139] was then calculated using only the portion of the predictive mean vector that was associated with the given missingness pattern. If no donors were available who had predicted means within a multivariate delta of the recipient's vector of predicted means, the neighborhood was abandoned, and the respondent with the closest Mahalanobis distance was selected as the donor. The procedure is described in detail in Appendix C.

10.4 Imputation-Revised Specific Health Insurance Variables (Constituent Variables Method, First Stage)

The constituent variables method of creating the final imputation-revised health insurance variables amounted to imputing missing values in each of the edited health insurance variables that, when combined together, constituted "overall health insurance." In the first stage of this method, which is described in this section, four imputation-revised specific health insurance variables were created representing 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). Missing values in these variables were imputed in a multivariate imputation. These final variables were derived from the edited variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN, respectively. The second stage is described in Section 10.5.

10.4.1 Order of Modeling Health Insurance Variables (Constituent Variables Method, First Stage)

The first step in imputing the four specific health insurance variables was to determine the order in which the variables were to be modeled. A motivation for using a hierarchy in PMN for drug use variables is provided in Appendix C; this same rationale was used in developing the hierarchy for the health insurance variables. For a model predicting whether a respondent had a specific type of health insurance, other types of health insurance were useful covariates. Following a provisional imputation of missing health insurance values, the indicators earlier in the sequence were used as covariates for health insurance variables later in the sequence. Any imputed values in the health insurance variables were considered temporary at this point. This was because the final imputation was not done for health insurance variables until the modeling was completed for all four specific health insurance variables. The following is the order in which the health insurance indicators were imputed: CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN.

10.4.2 Setup for Model Building (Constituent Variables Method, First Stage)

Once the hierarchy of health insurance variables was determined, the next step was to define respondents, nonrespondents, and the item response mechanism. For an individual to have been considered an item respondent for the specific health insurance variables, he or she had to have complete data for the four edited specific health insurance variables. Imputation for CAIDCHIP, CHAMPUS, and private health insurance were conducted within the four age

[139] See Appendix C for a definition of Mahalanobis distance. A definition also can be found in Manly (1986).

groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. Imputation for Medicare was conducted within the following three age groups: 12 to 17, 18 to 64, and 65 or older.[140]

Response propensity adjustments were then computed for each age group to make the item respondent weights representative of the entire sample. The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model included a centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three-level income variable. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups (i.e., 18 to 25, 26 to 64, and 65 or older), the additional predictors of marital status, education level, and employment status also were considered in each model.

10.4.3 Sequential Model Building (Constituent Variables Method, First Stage)

Starting with CAIDCHIP, the probability that an individual was covered by a given type of health insurance was modeled for item respondents, within each age group, using the nonresponse-adjusted weights. For the models, the parameters were estimated using logistic regression in SUDAAN®.[141] The predictors included in all models were centered age, centered age squared, gender, race/ethnicity, population density, and percentage of housing in that segment that was owner-occupied. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender except for the 65 years of age. For the three older age groups (i.e., 18 to 25, 26 to 64, and 65 or older), the additional predictors of marital status, education level, and employment status also were considered in each model. Additional predictors were specific to each model, depending upon the response variable of interest, and are listed below.

CAIDCHIP: household size; a four-level family income variable;[142] binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, or social security, and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.

MEDICARE: for respondents 18 or older, a binary indicator of whether the respondent was on social security; and for respondents younger than 18, a binary indicator of whether anyone in the respondent's family in the household received social security.

---

[140] The age groups 18 to 25 and 26 to 64 were combined for the Medicare variable 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 only Medicare if he or she was not working due to disability. This was true regardless of whether the respondent was 18 to 25 or 26 to 64 years old.

[141] SAS®-callable SUDAAN® was used to fit the binomial and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in the *SUDAAN® User's Manual, Release 9.0* (RTI, 2004). SAS® software is a registered trademark of SAS Institute, Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

[142] The four levels of the family income variable were less than $20,000; $20,000 to $49,999; $50,000 to $74,999, and $75,000 or more.

CHAMPUS: a binary indicator of whether the respondent (or, if the respondent was younger than 18, the respondent's family in the household) received income from sources other than those given in the binary income questions (see Chapter 9 for details); a three-level income variable;[143] and for respondents 18 or older, an indicator of whether the respondent had ever been in the military service, designated by an imputation-revised version of the edited variable SERVICE.[144]

PRVHLTIN: household size; a four-level family income variable (the same variable that was used in the CAIDCHIP model); binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, social security, or from sources other than those given in the binary income questions (see Chapter 9 for details); and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.[145]

The complete summary of the health insurance models can be found in Appendix F.

10.4.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods (Constituent Variables Method, First Stage)

Following the modeling for the four specific health insurance variables corresponding to CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN, in the sequence listed in Section 10.4.1, missing values were replaced by provisional imputed values. This was necessary so that these variables could have been used as covariates in subsequent models. Although no provisional imputed values were used to build the models, it was necessary to calculate predicted means for all respondents, including item nonrespondents, using the parameter estimates from the models. This sometimes required the use of the provisional values for the covariates. The predicted probabilities from these models were used to assign provisional values using the univariate predictive mean neighborhood (UPMN) imputation method as described in Appendix C.

10.4.5 Assignment of Provisional Imputed Values (Constituent Variables Method, First Stage)

Separate assignments of provisional values were performed within the age groups that were used for each of the respective first three health insurance variables.

10.4.6 Multivariate Imputation of the Specific Health Insurance Variables (Constituent Variables Method, First Stage)

The final imputed values for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN were obtained using neighborhoods built upon a vector of predicted means. For these four

---

[143] The three levels were less than $20,000, $20,000 to $49,000, and $50,000 or more.

[144] The variable SERVICE generally had a very low level of missingness (0 missing values in the 2005 survey). Because covariates in these models were not supposed to have any missing values, the missing value in the SERVICE variable was randomly imputed as a "yes" if the random number was greater than the mean value of SERVICE across all the other respondents, and "no" otherwise.

[145] If the respondent did not have other family members in the household, the family income binary indicators listed as predictors were equivalent to the personal income binary indicators.

variables, the imputation method used was the PMN procedure, as described in Appendix C. Similar to the response propensity models, the multivariate assignments were done separately within the same four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. No logical constraints were applied to the health insurance variables, since no internal inconsistencies would have resulted from any type of donor. However, a number of likeness constraints were applied, depending upon the missingness pattern. The variables that were included as likeness constraints were highly correlated with the response variables, but (in most cases) could not have been included as predictors in the models, due to the large number of missing values in the predictors. In general, any nonmissing values that the recipient had for CAIDCHIP, MEDICARE, CHAMPUS, or PRVHLTIN had to match between donor and recipient, though this constraint was often the first one that was loosened. In addition, the donor's predicted mean(s) for each variable that was missing was required to be within 5 percent of the recipient's predicted mean(s). This was usually the last constraint to be loosened. Finally, specific likeness constraints were associated with each of CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Constraints associated with each variable are discussed briefly below. The order in which the constraints were loosened depended upon the missingness pattern, and these constraints are described in detail in Appendix G. The portions of the full predictive mean vector used to create the multivariate neighborhoods for each missingness pattern, with accompanying adjustments, are provided in Appendix H.

CAIDCHIP

The donor and recipient had to have the same status regarding whether or not a respondent's family had received any government public assistance. This was measured by the variable GOVTPROG, which is described in Chapter 9.

MEDICARE

A respondent of working age (between the ages of 18 and 64) could have received Medicare only if he or she were not working due to disability. If MEDICARE was missing, a constraint was included that required donors and recipients to have had the same status in this regard, using the appropriate level of the variable JBSTATR (respondent work situation in the past week). This constraint was never loosened. In addition, the donor and recipient had to have the same status regarding whether or not a respondent's family had received social security.

CHAMPUS

In the models for CHAMPUS, two variables were included as covariates that also were used as likeness constraints. An imputation-revised version of the variable SERVICE (whether the respondent had ever been in the military service) was used in the CHAMPUS model, whereas SERVICE was used directly as a likeness constraint. The other variable was a binary indicator of whether the respondent (or the respondent's family in the household, if the respondent was younger than 18) received income from sources other than those given in the binary income questions (see Chapter 9 for details). Neither likeness constraint was loosened in the 2005 survey for any of the age groups, making their inclusion in the models unnecessary.

PRVHLTIN

In the model for PRVHLTIN, a four-level income variable was used as a covariate that also was used as a likeness constraint for the youngest three age groups. This likeness constraint was never loosened in the 2005 survey, making its inclusion in the models unnecessary for these three age groups. If it had been loosened, the donor and recipient would have been required to have the same value for a two-level income variable (less than $20,000 and $20,000 or more). For respondents 65 years of age or older, this two-level income variable was used as an initial likeness constraint and was never loosened in the 2005 survey.

10.5 Imputation-Revised Any-Other-Health-Insurance and Overall-Health-Insurance Recoded Variable (Constituent Variables Method, Second Stage)

The constituent variables method of creating the final imputation-revised health insurance variables amounted to imputing missing values in each of the edited health insurance variables that, when combined together, constituted "overall health insurance." In the second stage of this method, which is described in this section, a variable is created (IROTHHLT) that indicates 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. The final overall health insurance indicator is created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT.

10.5.1 Order of Modeling Health Insurance Variables (Constituent Variables Method, Second Stage)

Only one variable required imputation in the second stage. Therefore, an order of imputation was unnecessary.

10.5.2 Setup for Model Building (Constituent Variables Method, Second Stage)

Imputation for the any-other-health-insurance variable was conducted within the following age groups: 12 to 17, 18 to 25, and 26 or older.[146] For a respondent to have been considered an item respondent for modeling the any-other-health-insurance variable, he or she first had to have been part of the domain, which included respondents who had either a reported or imputed "no" value to all four imputation-revised specific health insurance variables (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT). Among respondents who were part of the domain, item respondents had to have complete data for the variable ANYOTHER, as defined in Section 10.2.2. Response propensity adjustments were computed within each age group to make the item respondent weights representative of the entire domain. The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model included a centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three-level income variable. There were also predictors

[146] Three age groups were used, instead of four, due to the small number of respondents who would have been included in the 65 or older age group.

that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the two older age groups (i.e., 18 to 25 and 26 or older), the additional predictors of marital status, education level, and employment status also were considered in each model.

10.5.3 Sequential Model Building (Constituent Variables Method, Second Stage)

The probability that an individual was covered by any other health insurance was modeled for item respondents within the domain defined in the previous section, within each age group, using the nonresponse-adjusted weights. The parameters were estimated using logistic regression in SUDAAN®, with the same base set of predictors that were used for the specific health insurance variables. In particular, these included centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three-level income variable. This base set also consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the two older age groups (i.e., 18 to 25 and 26 or older), the additional predictors of marital status, education level, and employment status also were considered in each model. Additional predictors were specific to the any-other-health-insurance model: household size, binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, social security, and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.[147]

The complete summary of the health insurance models can be found in Appendix F.

10.5.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods (Constituent Variables Method, Second Stage)

Following the modeling of the any other health insurance variable, missing values were replaced by imputed values. In the usual way, predicted means were calculated for all respondents, including item nonrespondents, using the parameter estimates from the models. The predicted probabilities from these models were used to assign imputed values using the UPMN imputation method as described in Appendix C.

10.5.5 Assignment of Imputed Values (Constituent Variables Method, Second Stage)

Separate assignments of provisional values were performed within the three age groups. The imputed values from these assignments were considered final. The imputation-revised version of the any other health insurance variable was called IROTHHLT.

[147] If the respondent did not have other family members in the household, the family income binary indicators listed as predictors were equivalent to the personal income binary indicators.

10.6 Creation of the Final Overall Health Insurance Variable (Constituent Variables Method)

The final overall health insurance indicator 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. This was recorded using the variable IRINSUR4, which was distinguished from the overall health insurance variable that was created using the old method, IRINSUR3. Though IRINSUR4 was technically a recoded variable created from other variables, an imputation indicator was nevertheless created, called IIINSUR4. Specifically, IIINSUR4 was set to "3" if any of the five constituent health insurance variables were imputed, "2" if none of the five variables were imputed and at least one was logically assigned, and "1" otherwise.

References

Chen, P., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2007). Person-level sampling weight calibration [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 3, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39; RTI/0209009). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

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/]

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.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376-382.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley & Sons.

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

Folsom, R. E., & Witt, M. B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. In *Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Toronto, Ontario, Canada* (pp. 428-433). Alexandria, VA: American Statistical Association.

Grau, E. A., Barnett-Walker, K., Copello, E., Frechtel, P., Licata, A., Liu, B., Martin, P., & Odom, D. M. (2005, March). Imputation report [2003]. In *2003 National Survey on Drug Use and Health: Methodological resource book* (Section 5, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 28, RTI/0208726). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k3]

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. (2004). *2002 National Survey on Drug Use and Health: Procedures for editing interviewer-administered data in the 2002 NSDUH computer-assisted interview* (for inclusion in the 2002 methodological resource book; report prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.595). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k2]

Kroutil, L. A., Handley, W., Suresh, P., Felts, B., & Bradshaw, M. (2007, February). Editing and coding report [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39; RTI/0209009.173.007). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

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

Manly, B. F. J. (1986). *Multivariate statistical methods: A primer*. London, England: Chapman and Hall.

Morton, K. B., Chromy, J. R., Hunter, S. R., & Martin, P. C. (2006, February). Sample design report [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 8, RTI/0209009). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

Office of Applied Studies, Substance Abuse and Mental Health Services Administration. (2001). *Development of computer-assisted interviewing procedures for the National Household Survey on Drug Abuse.* (Report No. DHHS Publication No. SMA 01-3514, Methodology Series M-3). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies. [Available at http://www.oas.samhsa.gov/nhsda/methods.cfm#Pre2k and http://www.oas.samhsa.gov/nhsda/CompAssistInterview/toc.htm]

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register, 62*(210), 58781-58790. [Available at http://www.whitehouse.gov/omb/fedreg/1997standards.html]

Penne, M. A., Lessler, J. T., Bieler, G., & Caspar, R. (1998). Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI field experiment. In *Proceedings of the 1998 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Dallas, TX* (pp. 744-749). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

RTI. (2004). *SUDAAN language manual: Release 9.0.* Research Triangle Park, NC: RTI International.

RTI. (2006). CAI specs for programming [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book*. (Prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39; RTI/0209009). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

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.

SAS Institute. (1999). *SAS/STAT user's guide: Version 8.* Cary, NC: SAS Institute.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). Boca Raton, FL: Chapman and Hall/CRC.

Shiffman, S., Hickcox, M., Gnys, M., Paty, J. A., & Kassel, J. D. (1995, March). *The Nicotine Dependence Syndrome Scale: Development of a new measure.* Poster presented at the annual meeting of the Society for Research on Nicotine and Tobacco, San Diego, CA.

Shiffman, S., Waters, A. J., & Hickcox, M. (2003). The Nicotine Dependence Syndrome Scale: A multi-dimensional measure of nicotine dependence. Unpublished manuscript.

Singh, A. C., & Mohl, C. A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology, 22*, 107-115.

Westlake, M., Aldworth, J., Barnett-Walker, K., Copello, E., Chen, P., Gordek, H., & Laufenberg, J. (2007). Questionnaire dwelling unit-level and person pair-level sampling weight calibration [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 4, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39; RTI/0209009). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

Williams, R. L., & Chromy, J. R. (1980). SAS sample selection MACROS. In *Proceedings of the Fifth International SAS Users Group International Conference* (pp. 382-396). Cary, NC: SAS Corporation.

Appendix A: Hot-Deck Method of Imputation

Appendix A: Hot-Deck Method of Imputation

A.1 Introduction

Typically, with the hot-deck method of imputation, missing responses for a particular variable (called the "base variable" in this Appendix) are replaced by values from similar respondents with respect to a number of covariates (called "auxiliary variables" in this Appendix). If "similarity" is defined in terms of a single predicted value from a model, these covariates 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 missing value is called the "donor."

For the 2005 National Survey on Drug Use and Health (NSDUH),[148] the imputation procedure used for most variables requiring imputation was the Predictive Mean Neighborhood method (PMN), which is a combination of predictive mean matching (Rubin, 1986) and unweighted random nearest neighbor hot deck (NNHD). No other type of hot-deck method was used to impute missing values in the 2005 survey. Although only one hot-deck imputation method was used in the 2004-2005 surveys, two other methods were used in past surveys. The three methods, which are each discussed in this document, are unweighted sequential hot deck, weighted sequential hot deck, and unweighted random NNHD. The first method, the unweighted sequential hot deck, was the exclusive method of hot-deck imputation used for the 1991 to 1998 surveys and the paper-and-pencil interviewing (PAPI) sample of the 1999 survey. This method was used for all demographic variables in the 1999 survey, but not used for other variables. In the 2000 survey, the unweighted sequential hot-deck method was used only for education and employment status variables and has not been used since the 2001 surveys. However, it remains in this appendix for historical purposes and for comparison with the other two methods. Starting in the 2002 survey, missing values in the immigrant variables required imputation. In the 2002 and 2003 surveys, the method used for these variables was the weighted sequential hot deck, which is also described in this report. More information on weighted sequential hot-deck imputation is available in Cox (1980, pp. 721-725). However, beginning with the 2004 survey, the immigrant variables were imputed using the PMN method, as explained in Chapter 5. Hence, the hot-deck method was applied only in the 2004 survey as a step within PMN, where the NNHD was used.

A step that is common to all hot-deck methods is the formation of imputation classes, which is discussed in Section A.2. This is followed by a general description of the three hot-deck methods as discussed in Sections A.3 to A.5. With each type of hot-deck imputation, the identities of the donors are generally tracked. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

---

[148] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

A.2 Formation of Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation procedures were implemented independently within classes defined by the cross of the auxiliary variables. These classes were defined by logical and likeness constraints, which are described in the main body of this report. Classes defined by the likeness constraints were collapsed if insufficient donors were available, and classes defined by logical constraints were not collapsed, due to the possibility of a resulting inconsistency with preexisting nonmissing values.

A.3 Unweighted Sequential Hot Deck

In the surveys where the unweighted sequential hot deck method was used, its implementation involved three basic steps. After the imputation classes were formed, the file was appropriately sorted and imputed values were assigned as described in the following sections.

A.3.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was 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 were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous surveys to sort the files prior to imputation:

- **Straight Sort**. A set of variables was sorted in ascending order by the first variable specified. Then, within each level of the first variable, the file was sorted in ascending order by the second variable specified, and so forth. For example

1	1	1
1	1	2
1	2	1
1	2	2
1	3	1
1	3	2
2	1	1
2	1	2
2	2	1
2	2	2
2	3	1
2	3	2

- **Serpentine Sort**. A set of variables was sorted so that the direction of the sort (ascending or descending) changed each time the value of a variable changed. For example

1	1	1
1	1	2

1	2	2
1	2	1
1	3	1
1	3	2
2	3	2
2	3	1
2	2	1
2	2	2
2	1	2
2	1	1

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.

A.3.2 Replacing Missing Values

The file was sorted and then read sequentially. Each time an item respondent was encountered (i.e., the base variable was nonmissing), the base variable response was stored, updating the donor response. Any subsequent nonrespondent in the file received the stored donor response, which in turn resulted in a statistically imputed response. A starting value was needed if an item nonrespondent was the first record in a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value. Because the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

A.3.3 Potential Problem

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there was the risk of several nonrespondents appearing next to one another on the sorted file. To detect this problem in NSDUH, the imputation donor was identified for every item being imputed. Then, when frequencies by imputation donor were examined, the problem was detected if several nonrespondents were aligned next to one another in the sort. When this problem occurred, sort variables were added or eliminated, or the order of the variables was rearranged.

A.4 Weighted Sequential Hot Deck

The steps taken to impute missing values in the weighted sequential hot-deck method were equivalent to those of the unweighted sequential hot deck. The details on the final imputation, however, differed with the incorporation of sampling weights. The first step, as always, was the formation of imputation classes. Afterwards, two additional steps, as described below, were implemented.

A.4.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was 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 were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous surveys to sort the files prior to imputation: straight sort and serpentine sort. Both of these methods are described in detail in Section A.3.1.

A.4.2 Replacing Missing Values

The procedure used in the 2005 survey followed directly from Cox (1980). Specifically, once the imputation classes are formed, the data is divided into two datasets: one for respondent and one for nonrespondents. Scaled weights $v(j)$ are then derived for all nonrespondents using the following formula:

$$v(j) = w(j)s(+)/w(+); \quad j = 1, 2, ..., n,$$

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 the all nonrespondents, and $s(+)$ is the sum of the sample weights for all the respondents (Cox, 1980). The respondent data file is partitioned into zones of width $v(j)$, where the imputed value for the j^{th} nonrespondent is 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, which could be implemented using the Chromy-Williams sample selection software (Williams & Chromy, 1980). Furthermore, Iannacchione (1982) revised the Chromy-Williams sample selection software so that each step of the weighted sequential hot deck is executed in one SAS macro run.

A.4.3 Benefits of Weighted Sequential Hot Deck

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there is the risk of several nonrespondents appearing next to one another in the sorted file. An imputed value could still be found for those cases, since the algorithm would select the previous respondent in the file. However, some modifications are required in the sorting procedure to prevent a single respondent from being the donor for several nonrespondents (see Section A.3.3). With the weighted sequential hot-deck method, on the other hand, this problem does not occur, because the weighted hot deck controls the number of times a donor can be selected. In addition, the weighted hot deck allows each respondent the chance to be a donor, since a respondent is selected within each $v(j)$.

The most important benefit of the weighted sequential hot-deck method, however, is the elimination of bias in the estimates of means and totals. This type of bias is particularly present when the response rate is low or the covariates explain only a small amount of variation in the specified variable. In addition, many surveys, besides NSDUH, sample subpopulations at different rates, and using the sample weights allows, in expectation, 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).

A.5 Unweighted Random Nearest Neighbor Hot Deck

As with the other methods, the unweighted random NNHD method was implemented in three steps. After the imputation classes were formed, a neighborhood of potential donors was created, from which imputed values were assigned, as described in the following sections.

A.5.1 Creating a Neighborhood of Potential Donors

First, a metric was defined to measure the distance between units, based on the values of the covariates. Then, a neighborhood was created of potential donors "close to" the recipient based on that metric. For example, the distance between the values of the recipient and potential donors for each of the auxiliary variables were calculated, and then the donors for the neighborhood were chosen such that the maximum of these distances was less than a certain value, referred to as "delta." This neighborhood was restricted, using the imputation classes defined above, so that the potential donors' values of the base variable were consistent with the recipient's preexisting nonmissing values of related variables. In NSDUH, the values of the auxiliary variables were represented by a predicted mean from a model so that the distance metric was a univariate Euclidean distance between the predicted mean of the recipient and the potential donors. The distance was relative when dividing this value by the predicted mean of the recipient, resulting in delta as a percentage.

In application, if the predicted means were probabilities, the values of delta varied depending upon the value of the predicted mean. In this case, each delta was defined as 5 percent of the predicted probability if the probability was less than 0.5 and was defined as 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed 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 A.1.

Table A.1 Values of Delta for Various Predicted Probabilities

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

A.5.2 Randomly Selecting a Donor for the Recipient from the Neighborhood of Donors

From the neighborhood of donors created in the previous step, a single donor was randomly selected. The base variable values for this single donor replaced those of the recipient. The selection was conducted as a simple random sample[149] because weights were incorporated in determining the neighborhood mean, which was the predicted mean. Alternatively, a weighted

[149] In the surveys prior to the 2005 NSDUH, this probability was incorrectly calculated. Instead of each donor in the neighborhood (of size *n*) being assigned a probability of $1/n$ of being selected, the first and last donors in the neighborhood were assigned a probability of $1/(2(n-1))$ of being selected, and the remaining donors were assigned a probability of $1/(n-1)$ of being selected. This probability was corrected in the 2005 survey.

selection could have been employed if weights had not been used to determine the neighborhood mean. If no donors were available with predicted means within delta of the recipient's predicted mean(s), the neighborhood was abandoned and the donor with the closest predicted mean(s) was chosen.[150] This probability was done to reduce the potential for bias.

[150] In the surveys prior to the 2006 NSDUH, this probability sometimes was incorrectly applied. In some cases, the neighborhood was not sorted by the predicted means, and in other cases, a donor was randomly selected. There were also some situations where the donor with the closest predicted means was chosen with the delta constraint in place. These procedures will be corrected in the 2006 survey.

Appendix B: Technical Details about the Generalized Exponential Model

Appendix B: Technical Details about the Generalized Exponential Model

B.1 Introduction

For the 2005 National Survey on Drug Use and Health (NSDUH),[151] as well as all of the surveys since the computer-assisted interview (CAI) was introduced in 1999, a special case of the generalized exponential model (GEM)[152] has been used for weighting procedures. This special case was known as the item response propensity model, where weights among item respondents were adjusted to account for the weights of the item nonrespondents. The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures. Additional technical details concerning the GEM are described in the following sections.

B.2 Distance Function

Let $\Delta(\mathbf{w},\mathbf{d})$ denote the (scalar) distance between the vector of initial weights $\mathbf{d}=(d_1,...,d_n)'$ and the vector of adjusted weights $\mathbf{w}=(w_1,...,w_n)'$, where *n* represents the size of the sample *s*. The distance function minimized under the GEM subject to calibration constraints is given by

$$\Delta(\mathbf{w},\mathbf{d})=\sum\nolimits_{k\in s}\frac{d_k}{A_k}\left\{(a_k-\ell_k)\log\left(\frac{a_k-\ell_k}{c_k-\ell_k}\right)+(u_k-a_k)\log\left(\frac{u_k-a_k}{u_k-c_k}\right)\right\}, \qquad \text{(B2.1)}$$

where $a_k=w_k/d_k, A_k=(u_k-\ell_k)/[(u_k-c_k)(c_k-\ell_k)]$ and ℓ_k,c_k,u_k are prescribed real numbers. Let $\mathbf{T}_x$ denote the *p*-vector of control totals corresponding to the vector of predictor variables $\mathbf{x}=(x_1,...,x_p)'$. Then the calibration constraints for the above minimization problem are

$$\sum\nolimits_{k\in s}d_k a_k\mathbf{x}_k=\mathbf{T}_x\,, \qquad \text{(B2.2)}$$

where $\mathbf{x}_k$ represents the vector of predictor variables corresponding to the k^{th} element of the sample. The solution of the above minimization problem, if it exists, is given by a GEM with model parameters $\boldsymbol{\lambda}$, as follows:

---

[151] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[152] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

$$a_k(\boldsymbol{\lambda}) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp\{A_k \mathbf{x}_k' \boldsymbol{\lambda}\}}{(u_k - c_k) + (c_k - \ell_k)\exp\{A_k \mathbf{x}_k' \boldsymbol{\lambda}\}}. \quad \text{(B2.3)}$$

Note that the number of parameters in GEM should be $\leq n$ (i.e., the size of the sample *s*). It follows from (B2.3) that

$$\ell_k < a_k < u_k,\ k = 1,..,n \quad \text{(B2.4)}$$

The usual Raking-ratio method (Singh & Mohl, 1996) of weight adjustment is a special case of GEM, noting that for $\ell_k = 0, u_k = \infty,\ c_k = 1,\ k = 1,..,n,$

$$\Delta(\mathbf{w},\mathbf{d}) = \sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k(a_k - 1)$$

and $a_k(\boldsymbol{\lambda}) = \exp(\mathbf{x}_k'\boldsymbol{\lambda})$.

The logit method of Deville and Särndal (1992) is also a special case of GEM, setting $\ell_k = \ell,\ u_k = u,\ c_k = 1$ for all *k*. The new method was introduced by Folsom and Singh (2000).

B.3 GEM Adjustments for Extreme Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (B2.3) can be justified for all three types of adjustment: extreme value treatment, nonresponse, and poststratification.

For extreme value treatment via winsorization, denote the winsorized weights by $\{b_k\}$, where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = \text{med}\{d_k\} \pm 3 *$ IQR if d_k is an extreme weight (where *med* represents the median, IQR represents the interquartile range, and the median and quartiles for the weights are defined with respect to a suitable design-based stratum).

For the nonresponse adjustment, the sample is divided as before into two parts: s^*, the nonextreme weight subsample, and s^{**}, the extreme weight subsample. For nonextreme weights, the following are set: $\ell_2 = 1,\ c_2 = \rho^{-1},\ u_2 = u > \rho^{-1}$, where ρ is the overall response propensity; and, for extreme weights with high weights, they are set as, $\ell_k = \ell_1 m_k,\ c_k = \rho^{-1} m_k,\ u_k = u_1 m_k$, where $m_k = b_k / d_k$ and $1 \leq \ell_1 < \rho^{-1} = c_1 < u_1$, are prescribed numbers. Similarly, for extreme weights with low weights, $\ell_k = \ell_3 m_k,\ c_k = \rho^{-1} m_k,\ u_k = u_3 m_k$, and $1 \leq \ell_3 < \rho^{-1} = c_3 < u_3$.

For the poststratification adjustment, the following weights are set: for nonextreme weights, $\ell_k = \ell_2,\ c_k = c_2 = 1,\ u_k = u_2$; for high extreme weights, $\ell_k = \ell_1 m_k,\ c_k = m_k,\ u_k = u_1 m_k$; and for low extreme weights, $\ell_k = \ell_3 m_k,\ c_k = m_k,\ u_k = u_3 m_k$. The extreme value adjustment is identical to the poststratification adjustment, except for tighter bounds on extreme weights resulting from the final poststratification.

Notice that GEM allows for the flexibility of specifying different bounds for different subsamples; in addition, the lower bound (in the case of nonresponse adjustments) can be made to equal 1 by choosing the center $c_k > 1$.

B.4 Newton-Raphson Steps

Let $\mathbf{X}$ denote the $n \times p$ matrix of predictor values, and for the ν^{th} iteration, define

$$\Phi_k^{(\nu)} = \begin{cases} 1, & \nu = 0 \\ (u_k - a_k^{(\nu)})(a_k^{(\nu)} - \ell_k) \big/ \big((u_k - c_k)(c_k - \ell_k)\big), & \nu > 0. \end{cases}$$

Then, $\mathbf{T}_{\Phi\nu}$ is defined as an $n \times n$ diagonal matrix, whose k^{th} diagonal element is $d_k \Phi_k^{(\nu)}$. Then, at the Newton-Raphson iteration ν, the value of the p-vector $\boldsymbol{\lambda}$ is adjusted as

$$\boldsymbol{\lambda}^{(\nu)} = \boldsymbol{\lambda}^{(\nu-1)} + \left(\mathbf{X}'\mathbf{T}_{\Phi,\nu-1}\mathbf{X}\right)^{-1} \left(\mathbf{T}_x - \hat{\mathbf{T}}_x^{(\nu-1)}\right), \qquad \text{(B4.1)}$$

where $\boldsymbol{\lambda}^{(0)} = 1$.

The convergence criterion is based on the Euclidean distance,

$$\left\| \mathbf{T}_x - \hat{\mathbf{T}}_x^{(\nu)} \right\| = \sqrt{\left(\mathbf{T}_x - \hat{\mathbf{T}}_x^{(\nu)}\right)' \left(\mathbf{T}_x - \hat{\mathbf{T}}_x^{(\nu)}\right)}\,.$$

At each iteration, it is checked to see whether it is decreasing or not. If not, a half-step is used in the iteration increment.

B.5 Scaled Constrained Exponential Model

In previous surveys, constrained exponential models (CEMs) were used for poststratification and scaled CEMs were used for nonresponse adjustments. The CEM refers to the logit model of Deville and Särndal (1992) in which lower and upper bounds do not vary with k (i.e., $\ell_k = \ell$, $u_k = u$, and $c_k = c = 1$ such that $\ell < 1 < u$). Thus, it is a special case of GEM. For the nonresponse adjustment, Folsom and Witt (1994) modified CEM estimating equations by a scaling factor (ρ^{-1}, the inverse of the overall response propensity) such that $1 < \rho^{-1} a_k < \rho^{-1} u$. This implies that by choosing ℓ in CEM as ρ, it ensures that the scaled adjustment factor for nonresponse is at least 1.

Appendix C: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

Appendix C: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

C.1 Introduction

Since the introduction of the computer-assisted interview (CAI) in 1999 for the National Survey on Drug Use and Health (NSDUH),[153] one imputation method has been used for most variables requiring imputation. This method is the predictive mean neighborhood (PMN) method, and it was developed to cater to the specific needs of NSDUH. This approach has been used since the 1999 survey[154] and can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, PMN incorporates predicted means from models and the assignment of imputed values using neighborhoods determined by those predicted means.

C.2 Overview

C.2.1 Predictive Mean Neighborhood Method: Derived from Combining Nearest Neighbor Hot Deck and Predictive Mean Matching

The PMN method is a combination of two commonly used imputation methods: a non-model-based hot deck (nearest neighbor) and a modification of the model-assisted predictive mean matching (PMM) method of Rubin (1986). The PMN method enhances the PMM method. Specifically, the PMN method can be applied to both discrete and continuous variables, either individually or jointly. The PMN method also enhances the nearest neighbor hot-deck (NNHD) method so that the distance function used to find neighbors is no longer ad hoc.

A commonly used imputation method is a random NNHD (Little & Rubin, 1987, p. 65). With this method, donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (the donor has complete data, and the recipient does not). A donor set deemed close to the recipient, with respect to a number of covariates, is used to select a donor at random. For NSDUH, the set of covariates typically included demographic variables, as well as some other nonmissing drug use variables. In the case of NSDUH, to further ensure that a donor matched the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with drug use, such as age categories, were often used to restrict the set of donors. Furthermore, other restrictions involving outcome variables were imposed on the neighborhood.

Note that in NNHD, unlike sequential hot deck, a distance function is used to define closeness between the recipient and a donor. So, there is less of a problem of sparseness of the

[153] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[154] After the 1999 survey, only a CAI sample was selected.

donor class, but the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

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

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

However, the choice of donor is nonrandom. This nonrandomness leads to bias in the estimators of means and totals. It also tends to skew the distribution of outcome values to the center. Furthermore, as mentioned earlier, the PMM method is not applicable to discrete variables, because the distance function between the recipient's predicted mean (which takes continuous values) and the donor's outcome value (which takes discrete values) is not well defined.

C.2.2 Univariate and Multivariate Applications of the Predictive Mean Neighborhood Method

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

The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be tedious in practice because of the computational problems 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 the probability distribution underlying survey data, none of the existing solutions takes the survey design into account. However, since the 1999 surveys, in the application of the multivariate predictive mean neighborhood (MPMN) method to the imputation procedures, a multivariate model has been fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the hierarchy have a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for the use of univariate techniques for fitting multivariate (but conditional) predicted means.

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

C.3 Outline and Description of Method

The procedure for implementing UPMN and MPMN in NSDUH entailed six steps. Steps 2 through 5, and sometimes Step 6, as described below, were cycled through each of the drugs and drug use measures in the order determined by Step 1. Steps 4 and 5 (Steps 4 through 6 when applicable) could have been considered a variant of a random NNHD.

C.3.1 Step 1: Definition of Hierarchy

The first step was to determine the order in which variables were modeled so that variables early in the hierarchy could have been used for modeling the conditional predicted mean (i.e., they have the potential to have been part of the set of covariates for variables later in the hierarchy). Note that usually not all variables in the hierarchy were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For example, in the drug modules in NSDUH, different drugs needed to have been modeled, with different measures of drug use for each drug. It was therefore necessary to determine the order in which the combination of drugs and drug use measures would have been handled. Using the sequence of variables determined by this step, the procedure involved cycling through Steps 2 through 5 and sometimes Step 6. In the application of the PMN to NSDUH, the order of imputation for drugs was determined by considering such factors as the level of stigma associated with the drugs, the level of "missingness" in the data (see Appendix H), and the degree to which one set of drugs could have been used as predictors for other drugs. The order of drugs was given by cigarettes, smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, crack, and heroin. The order of drug use measures imputed was determined based on the natural hierarchy of the variables: lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and age of first use.

For each variable, Steps 2 through 5 were followed for NSDUH.

C.3.2 Step 2: Setup for Model Building and Hot-Deck Assignment

For each model that was fitted, two groups were created: complete data respondents and incomplete data respondents (item respondents and item nonrespondents, respectively). Complete data respondents had complete data across the variables of interest, and incomplete data respondents encompassed the remainder of respondents. If the final assignment was multivariate, complete data respondents must have had complete data across all the variables in the multivariate response vector. Models were constructed using complete data respondents only.

C.3.3 Step 3: Sequential Hierarchical Modeling

The model was built using the complete data respondents only with weights adjusted for item nonresponse. For the drug modules in NSDUH, lifetime usage indicators were modeled first because all other drug use indicators depended on an indication of lifetime use or nonuse. Once the hierarchy of drugs for lifetime usage was determined, lifetime usage indicators for individual drugs were modeled in a sequential fashion. The sequence used for the remaining combinations of drugs and drug use measures depended on what covariates were desired in the models and what variables were considered part of a multivariate set.

C.3.4 Step 4: Computation of Predicted Means and Delta Neighborhoods

Once the model was fitted, the predicted means for item respondents and item nonrespondents were calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element within that vector was chosen so that each respondent had exactly one predictive mean value.[155] This predicted mean was the matching variable in a random NNHD. It could have come directly from the model, it could have been adjusted to account for the conditioning on the time period, or (if it was the predicted value based on a model with a transformed response variable) it could have been back-transformed to the original units.

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

The pool of donors was further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this

[155] Alternatively, a provisional MPMN method could have been performed by using the predicted probabilities from the polytomous model. If this method occurred, the final MPMN would have been built based on probabilities from the polytomous model, as well as predicted means for the other variables in the multivariate set. See Step 6 (Section C.3.6) for a description of the MPMN.

type of constraint, called a "logical constraint," was given by age at first crack use, which must not have been less than age at first cocaine use. Other constraints, called "likeness constraints," were placed on the pool of donors to make the attributes of the neighborhood as close to that of the recipient as possible. For example, for age at first use, the age of the donor and the age of the recipient were restricted to have been the same whenever possible, and the donor and recipient must have come from States with similar usage patterns. A small value of delta also could have been considered as a likeness constraint. Whenever insufficient donors were available to meet the likeness constraints, including the preset small value of delta, the constraints were loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors was available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor was found using a sequential hot deck, where matching was done on the predicted mean. (Even though weights would not have been used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process, because weighting would already have been incorporated in the calculation of the predicted mean.)

If many variables were imputed in a single multivariate imputation, it was advantageous to preserve, as much as possible, correlations between variables in the data. However, the more variables that were included in a multivariate set, the less likely it was that a neighborhood could have been used for the imputation within a given delta. Even though there were many advantages to using multivariate imputation, one disadvantage, in several instances, was not being able to find a neighborhood within the specified delta.

C.3.5 Step 5: Assignment of Imputed Values Using a Univariate Predictive Mean Neighborhood

Using a simple random draw from the neighborhood developed in Step 4, a donor was chosen for each item nonrespondent. If only one response variable was imputed, the assignment step was a simple replacement of a missing value by the value of the donor. It was possible, however, that a donated quantity was a function of the final imputed value. For example, for 12-month frequency of drug use, 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, where the "total period" could have ranged up to a year. In the assignment step, the donor's proportion of total period was multiplied by the recipient's maximum possible number of days in the year that he or she could have used the substance.

The assignment step was multivariate if several response variables were associated with a single predicted mean, provided more than one of those response variables was missing. In that case, all of the missing values were imputed using the same donor. If there was more than one response variable associated with a single predicted mean, but not all of them were missing, only the missing values were replaced by those of the donor. The resulting imputed values were provisional if a multivariate predictive mean vector was needed in a final multivariate imputation. Otherwise, these values were final.[156]

The variables requiring imputation were part of a multivariate set if a multivariate predictive mean vector was used to match donors and recipients in a final multivariate imputation. If the variables were part of a multivariate set, it was necessary to cycle through Steps 2 through 5 for each variable in the set, then proceed to Step 6 after completing Steps 2 through 5 for the last variable in the set. If the variables were not part of a multivariate set, then it was only necessary to go through Steps 2 through 5 once, and proceeding to Step 6 was unnecessary. After the completion of either Step 5 (if a univariate predicted mean was used) or Step 6 (if a multivariate predictive mean vector was used), the next variable in the hierarchy requiring imputation was processed by returning to Step 2.

C.3.6 Step 6: Determination of Multivariate Predictive Mean Neighborhood and Assignment of Imputed Values

With the MPMN method, the neighborhood was defined based on a vector of predicted means rather than from a single predicted mean as in the univariate case. This vector may have encompassed a subvector of predicted means from a single categorical model (as with a polytomous logit model), in addition to scalar predicted means from any number of models with continuous response variables. For each item nonrespondent, a distance was calculated between the elements of this vector of predicted means, where the observed values were missing, and the corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predicted means that were calculated on the basis of past year and past month users were further adjusted to account for the probability that a respondent was a past year user or a past month user. For example, in NSDUH, the full predictive mean vector for alcohol included the following elements:

1. *recency, past month*: *P*(past month alcohol user | lifetime alcohol user);
2. *recency, past year, not past month*: *P*(past year but not past month alcohol user | lifetime alcohol user);
3. *12-month frequency*: *P*(the respondent used alcohol on a given day in the past year | past year user of alcohol)**P*(past year user of alcohol | lifetime alcohol user);[157]

---

[156] If the variable was part of a multivariate set upon which the MPMN method was applied, and provisional values were not needed for subsequent models, Steps 4 (creation of delta neighborhood) and 5 could have been skipped.

[157] For the 12-month frequency, 30-day frequency, and 30-day binge frequency, the models were fitted using logits. These logits were converted to probabilities when creating the predictive mean vector. 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. This, of course, was not true. However, 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 a predicted mean to be made conditional on what was known.

4. *30-day frequency*: *P*(the respondent used alcohol on a given day in the past month | past month user of alcohol)***P*(past month alcohol user | lifetime alcohol user); and
5. *30-day binge frequency*: *P*(the respondent was a binge drinker on a given day in the past month | past month user)***P*(past month alcohol user | lifetime alcohol user).

The subset of elements used to determine a neighborhood for a particular item nonrespondent depended on the missingness pattern of that item nonrespondent.[158] Moreover, if partial information was available on the recency of use, the predicted means was adjusted to account for that knowledge. For example, if a particular item nonrespondent was known as a past year alcohol user and his 12-month frequency was known, the elements above for which differences would have been calculated would be element #1 conditioned on past year use and elements #4 and #5. That is,

P(*Past month alcohol user* | *Lifetime alcohol user*) ÷ *P*(*Past year alcohol user* | *Lifetime alcohol user*),

P(*Respondent used alcohol on a given day in the past month* | *Past month user of alcohol*)* *P*(*Past month alcohol user* | *Lifetime alcohol user*) ÷ *P*(*Past year alcohol user* | *Lifetime alcohol user*), and

P(*Respondent was a binge drinker on a given day in the past month* | *Past month user*)**P*(*Past month alcohol user* | *Lifetime alcohol user*) ÷ *P*(*Past year alcohol user* | *Lifetime alcohol user*).

A neighborhood resulting from this vector of distances was constrained by a multivariate preset delta, such that the distance associated with each element of the predictive mean vector had to be less than the preset delta associated with that element. From the donors that satisfied the multivariate delta condition, a single neighborhood was created by first converting the vector of differences into a scalar distance measure. This scalar distance measure, called the Mahalanobis distance, 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 (sub-)vector for a given item respondent, and $\boldsymbol{\mu}_{NR}$ is the predictive mean (sub-)vector for a given item nonrespondent. The matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the predicted means, calculated using the subvector of predicted means associated with each missingness pattern, using complete data respondents within each age group and (where applicable) State rank group. Note that the Mahalanobis distance measure is a generalization of the corresponding Euclidean distance measure $\sqrt{(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})'(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})}$; the Mahalanobis distance standardizes the Euclidean distance by the variance-covariance matrix,

[158] Alternatively, the entire predictive mean vector could have been used to determine the neighborhood, regardless of the missingness pattern. Because many respondents in the multivariate set were missing only one item in the set, imputation was accomplished using UPMN, which is computationally much faster.

which is appropriate for random variables that are correlated or have heterogeneous variances. Also note that since the square of the Mahalanobis distance is a monotone function of the distance itself, the additional step of taking the square root of the squared distance was not performed in practice. The Mahalanobis squared distance was calculated only for those respondents who met the multivariate delta constraint. The neighborhood was determined by selecting the *k* smallest Mahalanobis squared distances within this subset of item respondents for a given item nonrespondent.

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

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

C.4 Comparison of PMN with Other Available Imputation Methods

The PMN methodology addresses all of the shortcomings of the unweighted sequential hot-deck method:

- **Ability to use covariates to determine donors is far greater than in the hot deck.** As with other model-based techniques, using models allows more covariates to be incorporated, including measures of use of other drugs, in a systematic fashion, where weights can be incorporated without difficulty. However, like a hot deck, covariates not explicitly modeled can be used to restrict the set of donors using logical constraints. If there is particular interest in having donors and recipients with similar values of certain covariates, they can be used to restrict the set of donors using likeness constraints even if they are already in the model.
- **Relative importance of covariates is determined by standard estimating equation techniques.** In other words, there are objective criteria based on methodology, such

as regression, which quantify the relationship between a given covariate and the response variable, in the presence of other covariates. Thus, the response variable itself is indirectly used to determine donors.

- **Problem of sparse neighborhoods is considerably reduced, making it easier to implement restrictions on the donor set.** Because the distance function is defined as a continuous function of the predicted mean, it is possible to find donors arbitrarily close to the recipient. Thus, it is less likely to have the problem of sparse neighborhoods for hot decking. Moreover, having sufficient donors in the neighborhood allows for imposing extra constraints on the donor set, which would be difficult to incorporate directly in the model.

- **Sampling weights are easily incorporated in the models.** The weighted hot deck can be viewed as a special case of PMN.

- **Correlations across response variables are justified by making the imputation multivariate.**

- **Choice of donor can be made random by choosing delta large enough that the neighborhood is of a size greater than 1.** Under the assumption that the recipient and the candidate donors in the neighborhood have approximately equal means, the random selection allows the case where the error distribution with mean zero can be mimicked. This helps to avoid bias in estimating means and totals, variances of which can be estimated as in two-phase sampling or by suitable resampling methods.

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

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

Appendix D: Race and Hispanic/Latino Group Alpha Codes

Appendix D: Race and Hispanic/Latino Group Alpha Codes

D.1 Introduction

For the 2005 National Survey on Drug Use and Health (NSDUH),[159] it was not uncommon for a respondent to have felt 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. Therefore, the size of the dictionary file 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 geographic entity and a race in the other-specify response, such as "Japanese Peruvian," the geographic entity was ignored and the respondent was coded as "Japanese." The geographic entity was recorded only if no other information was available, either in the other-specify response or in the non-other-specify response. As discussed in Chapter 4, many respondents provided a race in the alpha-specify response to the Hispanic/Latino group question and vice versa, so responses to both questions were examined in the creation of each variable. This appendix summarizes the procedures that were implemented to assign race and Hispanic/Latino values to respondents based on alpha-specify responses.

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 done using only the information from the non-other-specify category (categories) and the other-specify information was ignored.

D.2 Race

In the 2005 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 also 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

[159] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

responses to these two questions were considered simultaneously. The race questions used in the 2005 survey were as follows:

QD05: Which of these groups describes you? Just give me the number or numbers from the card.

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? Just give me the number or numbers from the card.

1 Asian Indian
2 Chinese
3 Filipino
4 Japanese
5 Korean
6 Vietnamese
7 Other (Specify)

The Hispanic/Latino group question (QD04) question is discussed in Section D.3. It also has another-specify response, which was gleaned for race information whenever race information was not available from QD05 or QD05ASIA.

D.2.1 Race Alpha Responses

The four types of race other-specify responses are listed and described in Chapter 4. Abbreviated descriptions are repeated here for convenience.

1. Directly Mapped Codes

Directly mapped codes were codes mapped to one or more of the categories given in the questionnaire (see Chapter 4). 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 "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, since the Polish census data indicated that nearly all Poles were white.

2. 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 eleven given in the questionnaire (see Chapter 4) 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, since 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).

3. Codes Informative for Formal Imputation Procedures

Some other-specify responses did not lead to definitive information about the respondent's race. However, the responses were used to limit the final imputation. With these codes, the final imputation was restricted according to the information that was available. No imputation was required, of course, 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.

4. 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 D.1 lists all the race codes used in the 2005 survey, along with supplementary information related to race codes. Special situations associated with the four types are discussed in the following sections. For most codes, the final assignment depended upon whether the response was given in QD05 or QD05ASIA. For codes described in #3 given 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 and the imputation indicator for the Hispanic/Latino indicator (IIHOIND) was set to 2 to indicate "logically assigned." See Chapter 4. Note that EDHOIND also could have been edited to a "no." This is discussed in the section on Hispanic/Latino codes (Section D.3.1.).

D.2.1.1 Handling of Directly Mapped Codes

For codes that were directly mapped, the final column of Table D.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/Alaskan Native" if in QD05 and "Asian Indian" if 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 D.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 peoples in the Polish population.

D.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, as explained in Chapter 4, the race was assigned by comparing a randomly generated number to the proportion of each racial category in the 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 common 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 (the randomly generated number rarely made it necessary for "other" to be selected) are identified with a superscript I in the third column of Table D.1. Rather than an "other" indication, the census sometimes gave general information ("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, first the race for each entity was determined (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 geographic entity. As stated earlier, in most cases the geographic 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, occasionally it was made clear that 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, 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" (or "dbl. census") is parenthetically indicated in the third column of Table D.1. If three indirectly mapped codes were indicated by the respondent, "triple census" is indicated. [160]

Details about how to handle census information for each indirectly mapped code are shown in Table D.2. Note that the race categories for each country listed in Table D.2 have been modified to conform to the race categories specified by the questionnaire. For example, the black race category from other countries has been modified to the black/African American category. Every category/restricted imputation level with a nonzero probability of selection is listed. If a code had an indirect map (using census data) for QD05, but a direct map for QD05ASIA, this is not specified in Table D.2. Instead, this information must be obtained from Table D.1. Explanations of the categories that are not self-explanatory are listed below.

a) "White or Mestizo": imputation was restricted to respondents who were either white or Mestizo (i.e., white and American Indian/Alaskan Native only). See Chapter 4 for the explanation of level 18 of EDRACE.
b) "Not American Indian": imputation was restricted to respondents who were of 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 4 for the explanation of level 19 of EDRACE.
c) "Multiple": imputation was restricted to respondents who were of multiple race. See Chapter 4 for the explanation of level 16 of EDRACE.

[160] 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 have been to impute race information for each component country, even if the "other" category was selected at random for one or more of them. See the entry for "Costa Rica" (code 78) in Table D.2 for an example.

D.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 said "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.[161] 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 was used along with the census data from the second country listed. More details about how specific restricted imputations are conducted are shown in Table D.3.

If the code was observed in QD05ASIA, then the imputation was not restricted only by what was written in the other-specify response, but it also was restricted only to 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.

D.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.

D.3 Hispanicity

As with the race questions, Hispanic/Latino respondents[162] 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? Just give me the number or numbers from the card.

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican
3 Central or South American
4 Cuban/Cuban American

---

[161] Due to an error, respondents who said only "Spanish" or only "Dominican" were mishandled in the 2004 survey and were treated the same way they were treated in the 2003 survey. Those who said "Spanish" only were mapped to "White," and those who said "Dominican" only were handled using census data for the Dominican Republic. In the 2005 survey, restricted imputation was used for these respondents.

[162] 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?"

5 Dominican (from Dominican Republic)
6 Spanish (from Spain)
7 Other (Specify)

Levels 5 and 6 were added to QD04 after the 2004 survey. They were included because, in previous years, there were a large number of other-specify responses for these categories.

The QD05 and QD05ASIA questions are discussed in Section D.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.

D.3.1 Hispanic/Latino Group Alpha Responses

There were only two types of Hispanic/Latino group other-specify responses: those that mapped to one or more EDQD04*xx* [163] variables and those that were ignored. There were no census-based routines and no responses that caused the imputation to be restricted. The imputation of Hispanic/Latino group was restricted only when race information was available.

Table D.4 lists all the Hispanic/Latino group codes used in the 2005 survey and the Hispanic/Latino groups to which they mapped. Note that these mappings utilized the arbitrary priority rule, provided in Chapter 4, that is used to create EDHOGRP, skipping the intermediate step of recording the Hispanic/Latino groups that were indicated in QD04. These are recorded in the EDQD04*xx* variables, which are described in Chapter 4. The creation of EDHOGRP is also described in Chapter 4. The Hispanic/Latino code 600, "Stated Clearly as Not Hispanic/Latino," was unique in that it could have been used to edit the Hispanic/Latino indicator, if needed. If QD03 was missing or 1, EDHOIND was edited to a 2 if this code appeared in QD04, QD05 or QD05ASIA (see Chapter 4).

Table D.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

[163] See Chapter 4 for a discussion of EDQD04*xx*.

Table D.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
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
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.

Table D.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
66	Bahamas & Haiti	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (dbl. census[3]) [I A]	QD05ASIA: O.A.
67	Brazil & Portugal	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (dbl. census[3]) [I A]	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.
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

Table D.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
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
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) [1]	Not a Direct Map
129	Qatar	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) [1]	Not a Direct Map

Table D.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
130	Bahrain	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) [I A]	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
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

Table D.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
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.
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) [I A]	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.

Table D.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
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) (dbl. census) [1]	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
203	White & American Indian/Alaska Native (incl. 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

Table D.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
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
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 (non-specific)	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 D.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 (non-specific)	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 D.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
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

Table D.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
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
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

Table D.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
411	White, Black/African American, Asian (non-specific)	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
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 (non-specific)	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

Table D.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
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) (dbl. census)[1]	QD05ASIA: O.A.
601	Mexican & El Salvadoran	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
602	Mexican & Honduran	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
603	Mexican & Nicaraguan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
604	Mexican & Costa Rican	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
605	Mexican & Panamanian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
606	Mexican & Colombian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
607	Mexican & Venezuelan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
608	Mexican & Ecuadorian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
609	Mexican & Peruvian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
610	Mexican & Bolivian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
611	Mexican & Chilean	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
612	Mexican & Argentine	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
613	Mexican & Paraguayan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
614	Mexican & Uruguayan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
615	Mexican & Brazilian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
616	Puerto Rican & Guatemalan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
617	Puerto Rican & El Salvadoran	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
618	Puerto Rican & Honduran	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
619	Puerto Rican & Nicaraguan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
620	Puerto Rican & Costa Rican	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
621	Puerto Rican & Panamanian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
622	Puerto Rican & Colombian	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.
623	Puerto Rican & Venezuelan	Indirectly mapped (QD05) (dbl. census)[1]	QD05ASIA: O.A.

Table D.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
624	Puerto Rican & Ecuadorian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
625	Puerto Rican & Peruvian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
626	Puerto Rican & Bolivian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
627	Puerto Rican & Chilean	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
628	Puerto Rican & Argentine	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
629	Puerto Rican & Paraguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
630	Puerto Rican & Uruguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
631	Puerto Rican & Brazilian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
632	Cuban & Guatemalan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
633	Cuban & El Salvadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
634	Cuban & Honduran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
635	Cuban & Nicaraguan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
636	Cuban & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
637	Cuban & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
638	Cuban & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
639	Cuban & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
640	Cuban & Ecuadorian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
641	Cuban & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
642	Cuban & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
643	Cuban & Chilean	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
644	Cuban & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
645	Cuban & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
646	Cuban & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
647	Cuban & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
648	Dominican & Guatemalan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
649	Dominican & El Salvadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
650	Dominican & Honduran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
651	Dominican & Nicaraguan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
652	Dominican & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
653	Dominican & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
654	Dominican & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
655	Dominican & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
656	Dominican & Ecuadorian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
657	Dominican & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
658	Dominican & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
659	Dominican & Chilean	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
660	Dominican & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
661	Dominican & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
662	Dominican & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
663	Dominican & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.

Table D.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
664	Spanish (from Spain) & Guatemalan	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
665	Spanish (from Spain) & El Salvadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
666	Spanish (from Spain) & Honduran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
667	Spanish (from Spain) & Nicaraguan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
668	Spanish (from Spain) & Costa Rican	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
669	Spanish (from Spain) & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
670	Spanish (from Spain) & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
671	Spanish (from Spain) & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
672	Spanish (from Spain) & Ecuadorian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
673	Spanish (from Spain) & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
674	Spanish (from Spain) & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
675	Spanish (from Spain) & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
676	Spanish (from Spain) & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
677	Spanish (from Spain) & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
678	Spanish (from Spain) & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
679	Spanish (from Spain) & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
680	Colombian & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
681	Colombian & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
682	Colombian & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
683	Colombian & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
684	Colombian & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
685	Colombian & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
686	Colombian & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
687	Colombian & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
688	Colombian & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
689	Venezuelan & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
690	Venezuelan & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
691	Venezuelan & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.

Table D.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
692	Venezuelan & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
693	Venezuelan & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
694	Venezuelan & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
695	Venezuelan & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
696	Venezuelan & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
697	Ecuadoran & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
698	Ecuadoran & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
699	Ecuadoran & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
700	Ecuadoran & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
701	Ecuadoran & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
702	Ecuadoran & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
703	Ecuadoran & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
704	Peruvian & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
705	Peruvian & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
706	Peruvian & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
707	Peruvian & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
708	Peruvian & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
709	Peruvian & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
710	Bolivian & Chilean	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
711	Bolivian & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
712	Bolivian & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
713	Bolivian & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
714	Bolivian & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
715	Chilean & Argentine	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
716	Chilean & Paraguayan	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
717	Chilean & Uruguayan	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
718	Chilean & Brazilian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
719	Argentine & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
720	Argentine & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
721	Argentine & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
722	Paraguayan & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
723	Paraguayan & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
724	Uruguayan & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
725	Guatemalan & Colombian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
726	El Salvadoran & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
727	Honduran & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
728	Nicaraguan & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
729	Costa Rican & Colombian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
730	Panamanian & Colombian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
731	Guatemalan & Venezuelan	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
732	El Salvadoran & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.

Table D.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
733	Honduran & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
734	Nicaraguan & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
735	Costa Rican & Venezuelan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
736	Panamanian & Venezuelan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
737	Guatemalan & Ecuadoran	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
738	El Salvadoran & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
739	Honduran & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
740	Nicaraguan & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
741	Costa Rican & Ecuadoran	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
742	Panamanian & Ecuadoran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
743	Guatemalan & Peruvian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
744	El Salvadoran & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
745	Honduran & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
746	Nicaraguan & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
747	Costa Rican & Peruvian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
748	Panamanian & Peruvian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
749	Guatemalan & Bolivian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
750	El Salvadoran & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
751	Honduran & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
752	Nicaraguan & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
753	Costa Rican & Bolivian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
754	Panamanian & Bolivian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
755	Guatemalan & Chilean	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
756	El Salvadoran & Chilean	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
757	Honduran & Chilean	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
758	Nicaraguan & Chilean	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
759	Costa Rican & Chilean	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
760	Panamanian & Chilean	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
761	Guatemalan & Argentine	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
762	El Salvadoran & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
763	Honduran & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
764	Nicaraguan & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
765	Costa Rican & Argentine	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
766	Panamanian & Argentine	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
767	Guatemalan & Paraguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
768	El Salvadoran & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
769	Honduran & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
770	Nicaraguan & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
771	Costa Rican & Paraguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
772	Panamanian & Paraguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
773	Guatemalan & Uruguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.

Table D.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
774	El Salvadoran & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
775	Honduran & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
776	Nicaraguan & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
777	Costa Rican & Uruguayan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
778	Panamanian & Uruguayan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
779	Guatemalan & Brazilian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
780	El Salvadoran & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
781	Honduran & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
782	Nicaraguan & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
783	Costa Rican & Brazilian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
784	Panamanian & Brazilian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
785	Guatemalan & El Salvadoran	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
786	Guatemalan & Honduran	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
787	Guatemalan & Nicaraguan	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
788	Guatemalan & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
789	Guatemalan & Panamanian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
790	El Salvadoran & Honduran	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
791	El Salvadoran & Nicaraguan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
792	El Salvadoran & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
793	El Salvadoran & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
794	Honduran & Nicaraguan	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
795	Honduran & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
796	Honduran & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
797	Nicaraguan & Costa Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
798	Nicaraguan & Panamanian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
799	Costa Rican & Panamanian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
800	Mexican & Jamaican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
801	Puerto Rican & Jamaican	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
802	Cuban & Jamaican	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
803	Dominican & Jamaican	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
804	Spanish (from Spain) & Jamaican	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
805	Mexican & European (not Spanish)	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
806	Puerto Rican & European (not Spanish)	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
807	Cuban & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.

Table D.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
808	Dominican & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
809	Spanish (from Spain) & Other European	Directly mapped (geographic category)	QD05: White
810	Trinidadian & Mexican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
811	Trinidadian & Puerto Rican	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
812	Trinidadian & Cuban	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
813	Trinidadian & Dominican	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
814	Trinidadian & Spanish (from Spain)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
815	Guatemalan & European (not Spanish)	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
816	El Salvador & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
817	Honduran & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
818	Nicaraguan & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
819	Costa Rican & European (not Spanish)	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
820	Panamanian & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
821	Colombian & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
822	Venezuelan & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
823	Ecuadoran & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
824	Peruvian & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
825	Bolivian & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
826	Chilean & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
827	Argentine & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
828	Paraguay & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
829	Uruguayan & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
830	Brazil & European (not Spanish)	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
831	(part) Mexican, ½ (part) white	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.

Table D.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
832	(part) Mexican, ½ (part) black	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
833	(part) Mexican, ½ (part) American Indian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
834	(part) Mexican, ½ (part) Hawaiian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
835	(part) Mexican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
836	(part) Mexican, ½ (part) Asian Indian	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
837	(part) Mexican, ½ (part) Chinese	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
838	(part) Mexican, ½ (part) Filipino	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
839	(part) Mexican, ½ (part) Japanese	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
840	(part) Mexican, ½ (part) Korean	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
841	(part) Mexican, ½ (part) Vietnamese	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
842	(part) Mexican, ½ (part) Other Asian	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
843	(part) Puerto Rican, ½ (part) White	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
844	(part) Puerto Rican, ½ (part) Black/African American	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
845	(part) Puerto Rican, ½ (part) American Indian/Alaska Native	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
846	(part) Puerto Rican, ½ (part) Native Hawaiian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
847	(part) Puerto Rican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
848	(part) Puerto Rican, ½ (part) Asian Indian	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
849	(part) Puerto Rican, ½ (part) Chinese	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
850	(part) Puerto Rican, ½ (part) Filipino	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
851	(part) Puerto Rican, ½ (part) Japanese	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map
852	(part) Puerto Rican, ½ (part) Korean	Indirectly mapped (QD05) (dbl. census) [I]	Not a Direct Map

Table D.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
853	(part) Puerto Rican, ½ (part) Vietnamese	Indirectly mapped (QD05) (dbl. census)[I]	Not a Direct Map
854	(part) Puerto Rican, ½ (part) Other Asian	Indirectly mapped (QD05) (dbl. census)[I]	Not a Direct Map
855*	(part) Hispanic/Latino, ½ (part) White	Directly mapped (racial category)	White
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) (dbl. census)	QD05ASIA: O.A.
871	Honduran & Haitian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
872	Guatemalan & Iranian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
873	Panamanian & Jamaican	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
874	Cuban & Thai	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
875	Venezuelan & Trinidadian	Indirectly mapped (QD05) (dbl. census)	QD05ASIA: O.A.
876	Puerto Rican & Arab	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
877	Puerto Rican & Virgin Islander	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
878	Mexican & Samoan	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
879	Salvadoran & Egyptian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
880	Costa Rican & Haitian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
881	Mexican & Iranian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.
882	Spanish & Barbadian	Indirectly mapped (QD05) (dbl. census)[I]	QD05ASIA: O.A.

Table D.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
883	Peruvian & Other Middle Eastern	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
884	Puerto Rican & African	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
885	Jamacian & Egyptian	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
886	Argentine & Turkish	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
890	Argentine, Cuban, & Spanish	Indirectly mapped (QD05) (triple census)	QD05ASIA: O.A.
891	Mexican, Cuban, & France	Indirectly mapped (QD05) (triple 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
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	Non-white non-specific/brown	Codes informative for formal imputation procedures	Not a Direct Map
905*	Hispanic/Latino non-white (incl. 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

Table D.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
917	Central/South American & Puerto Rican	Codes informative for formal imputation procedures	Not a Direct Map
918	Central/South American & Cuban	Codes informative for formal imputation procedures	Not a Direct Map
919	Central/South American & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
920	Central/South American & Spanish	Codes informative for formal imputation procedures	Not a Direct Map
921	Arab/Asian	QD05: Directly mapped (racial category) QD05ASIA: Directly mapped (geographic category)	QD05: White & Asian QD05ASIA: Other Asian
922	Arab/European	Directly mapped (geographic category)	QD05: White QD05ASIA: White & Asian
923	Arab/African	Directly mapped (geographic category)	QD05: White & Black/African American QD05ASIA: Asian & Black
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) (dbl. census) [I]	QD05ASIA: O.A.
927	West Indies & Cape Verde	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
928	Arab & Cape Verde	Indirectly mapped (QD05) (dbl. census) [I]	QD05ASIA: O.A.
929	Aryan	Directly mapped (geographic category)	QD05: White QD05ASIA: Asian Indian
930	Turkish & Lebanese	Directly mapped (racial category)	White
932	Puerto Rican & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
933	Puerto Rican & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
934	Cuban & Dominican	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 D.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
985	Bad data	Noninformative Code	Not a Direct Map
994	Unknown/"Don't Know"	Noninformative Code	Not a Direct Map
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 D.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 D.2.1.2 for details.

[2]Other Asian

[3]"dbl. census" is equivalent to "double census." See Section D.2.1.2 for details.

Table D.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
70	Mexico	9.3% White, 30.3% American Indian/Alaska Native, 60.3% White and American Indian/Alaska Native

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

Race Code	Race Name	Probabilities
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
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

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

Race Code	Race Name	Probabilities
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
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

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

Race Code	Race Name	Probabilities
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 D.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

Table D.3 Procedures for Restricted Imputation for Codes Informative for Formal Imputation Procedures (continued)

Race Code	Race Name	Restriction on Donors in Formal Imputation
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]
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
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
934	Cuban & Dominican	Donors were Cuban, Dominican, or both
935	Cuban & Spanish (from Spain)	Donors were Cuban, Spanish, or both
936	Dominican & Spanish (from Spain)	Donors were Dominican, Spanish, or both
904	Non-white nonspecific/brown	Donors were any race but single-race white
905	Hispanic/Latino non-white (incl. 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

[1]Even though a recipient may not have been 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]Since 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 D.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
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

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

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
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
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"

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

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
800	Non-Hispanic/Latino Country	Other Hispanic/Latino
801	Race 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 D.1.

Appendix E: Creation of Models Used to Allocate a Single Race among Multiple-Race Respondents

Appendix E: Creation of Models Used to Allocate a Single Race among Multiple-Race Respondents

E.1 Introduction

The race question QD06 appeared in the National Survey on Drug Use and Health (NSDUH)[164] from 1999 to 2002. Below is QD06 as it was presented in the 2002 survey.

QD06: Which **one** of these groups, that is [races chosen in QD05 and QD05ASIA], **best** 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 Indian
7 Chinese
8 Japanese
9 Filipino
10 Korean
11 Vietnamese
12 [Other from QD05ASIA, if applicable]
13 [Other from QD05, if applicable]
14 None of these

This question was presented to any respondent who selected more than one race category in questions QD05 and QD05ASIA combined. It was eliminated from the surveys after 2002, as per instructions from the Office of Management and Budget (OMB).

In the 1999-2002 surveys, multiple race respondents were each mapped to a single race using the response from QD06. This information was summarized in the variable IRRACE, which had four levels: American Indian/Alaska Native,[165] Asian/Other Pacific Islander,[166] black/African American, and white. However there was a high level of item nonresponse for QD06 in the 1999-2002 surveys. For multiple-race respondents who did not answer QD06, a single race was assigned using an arbitrary priority rule: black/African American, Asian/Other

[164] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[165] Alaska shall henceforth be abbreviated as AK in this appendix.

[166] Asian/Other Pacific Islander included respondents who gave at least one of the categories 4 through 11, and in most cases 12, in QD06. In some cases, it also included respondents who gave category 13 as a response, depending on the other-specify response.

Pacific Islander, American Indian/AK Native, and white.[167] Because the question QD06 was not presented in the surveys after the 2002 NSDUH, the variable IRRACE could not be created. To promote consistency with previous surveys, it was necessary to produce a variable similar to IRRACE in the 2005 NSDUH that distributed multiple race respondents among the four major race categories. The path chosen to assign a single race to a given multiple-race respondent in the 2005 NSDUH was to "simulate" QD06 using true QD06 responses from the 2000-2002 surveys.[168] Individual constituent races among multiple-race respondents were tracked using the variable EDRACE, which was described in Chapter 4. The levels of EDRACE are reproduced here in Table E.1 for convenience, where the races provided in the "Description" columns are the only races mentioned by the respondent.

Table E.1 Levels of EDRACE

Level	Description	Level	Description
1	White	**11**	White, Black/African American, American Indian/AK Native
2	Black/African American	**12**	White, Black/African American, Asian/Other Pacific Islander
3	American Indian/AK Native	**13**	White, American Indian/AK Native, Asian/Other Pacific Islander
4	Asian/Other Pacific Islander	**14**	Black/African American, American Indian/AK Native, Asian/Other Pacific Islander
5	White, Black/African American	**15**	White, Black/African American, American Indian/AK Native, Asian/Other Pacific Islander
6	White, American Indian/AK Native	**16**	More than one race, races unknown
7	White, Asian/Other Pacific Islander	**17**	Not white, races unknown
8	Black/African American, American Indian/AK Native	**18**	Either white or white and American Indian/AK Native
9	Black/African American, Asian/Other Pacific Islander	**19**	Not American Indian/AK Native, races unknown
10	American Indian/AK Native, Asian/Other Pacific Islander	**20**	Response of "Mexican" in QD05 but QD03 response was "not Hispanic/Latino"

An algorithm was used to take respondents with values of EDRACE between 5 and 15 (inclusive) and assign a single race corresponding to the values of EDRACE of 1, 2, 3, or 4. (For values of EDRACE of 16 or greater, a single race was determined using the formal imputation procedures described in Chapter 4.) The algorithm was completed in five steps.

[167] If black/African American was mentioned, the respondent was considered black, regardless of the other races mentioned. Otherwise, if Asian/Other Pacific Islander was mentioned, the respondent was Asian/Other Pacific Islander. This logic followed for the other races.

[168] Differences in the makeup of the questionnaire in the 1999 survey made it simpler to limit attention to results from the 2000-2002 surveys.

Step 1: Race Editing: Using the same editing rules that applied in the 2005 survey (i.e., the rules described in Chapter 4), race information for all respondents from the 2000-2002 surveys was edited. All respondents, except those who selected more than one race and whose constituent races were known, were discarded (i.e., kept only the multiple race respondents with $5 \leq EDRACE \leq 15$). The variable MULTRACECAT, which indicated the races that were selected (EDRACE – 4), was created. This variable had 11 levels.

Step 2: Creation of QD06RACE: For each level of MULTRACECAT, it was determined which major race category was selected in QD06. If none was selected or if QD06 was not encountered, the respondent was treated as an item nonrespondent.

Step 3: Adjusting Weights for Item Nonresponse: Weights for item nonresponse within the 11 levels of MULTRACECAT, within each survey year, were adjusted. So, 33 weight adjustments were done: 11 levels of MULTRACECAT within each of 3 survey years (2000, 2001, and 2002).

Step 4: Fitting of Logistic Regression Models: Predictive mean models for each level of MULTRACECAT were fit with data pooled across the survey years (2000-2002). So, 11 predictive mean models were fit and the parameter estimates were saved.

Step 5: Final Assignment of Single Race: The parameter estimates from Step 4 and the values of the covariates for each 2005 NSDUH multiple race respondent were used to estimate the probability that he or she would have chosen each source race, had he or she been asked QD06. A "best" race based on these predicted probabilities was randomly assigned.

Each of these five steps is described in detail below.

E.2 Steps Involved with the Algorithm

E.2.1 Race Editing

See Chapter 4 for a full description of the race/Hispanicity editing used in the 2005 survey. The only differences between the editing algorithm used in this appendix and the race/Hispanicity editing described in Chapter 4 are the following:

1) No editing of Hispanicity was done.
2) Records for respondents who did not have $5 \leq EDRACE \leq 15$ were discarded (this excluded single-race respondents or respondents who were considered multiple race, but for whom not all of their constituent races were known).

MULTRACECAT was assigned the value EDRACE – 4.

E.2.2 Creation of QD06RACE

The variable EDITQD06 was the edited version of QD06 and was used in the 1999-2002 surveys to determine what the respondent considered his or her "best" race. (See Kroutil, 2004, for details on the creation of the EDITQD06 variable in the 2002 survey.) If EDITQD06 was equal to one of the given race categories, the editing was straightforward. If it was missing, the respondent was treated as an item nonrespondent. All cases in which the "best" race was an other-specify response were examined carefully, and the major race category most similar to the one mentioned in the other-specify response was treated as the QD06 response. The variable QD06RACE was a condensed version of EDITQD06, with the four categories of IRRACE given earlier: 1 (white), 2 (black/African American), 3 (American Indian/AK Native), 4 (Asian/Other Pacific Islander), or missing.

There were about 50 cases for all three survey years (2000-2002) combined, for which other-specify responses were examined. In practically all these cases, the major race category was clear. For example, if a respondent selected "black/African American" in QD05 and "Norwegian" was written as an other-specify response, and if he or she chose "Norwegian" in QD06 as the best race, then the respondent was considered both white and black/African American with EDRACE = 5 and MULTRACECAT = 1. Since "Norwegian" was a directly mapped geographic category code[169] that mapped to white, QD06RACE would have been set to 1. The results of the first two steps are shown in Table E.2.

E.2.3 Adjusting Weights for Item Nonresponse

An interview respondent was considered an item nonrespondent for QD06 if his or her value for QD06RACE was missing. The weights of the item nonrespondents were reallocated to the item respondents using item response propensity models. One model was fit for each of the levels of MULTRACECAT within each of the survey years 2000-2002, for a total of 33 models.[170] The item response propensity model is a special case of the generalized exponential model (GEM),[171] which is described in greater detail in Appendix B. The starting list of covariates for each of the three item response propensity models consisted of age, census region, indicator of whether the householder was non-Hispanic/Latino black/African American, percentage black/African American population, percentage American Indian/AK Native population, percentage Asian population, and percentage of owner-occupied households. Some covariates were dropped due to convergence problems. The final set of covariates for each item response propensity model is shown in Table E.3.

E.2.4 Fitting of Logistic Regression Models

After the weights were adjusted for item nonresponse, logistic regression models were fit for each level of MULTRACECAT pooled across the three survey years of data (2000-2002 NSDUH), for a total of 11 models. Levels 1 through 6 of MULTRACECAT were combinations of two races, so these six models were dichotomous logistic regression models. Levels 7 through 10 of MULTRACECAT were combinations of three races, and level 11 was a combination of four races. The five models associated with levels 7 through 11 were polytomous logistic regression models.

---

[169] See Chapter 4 for a definition of "geographic category code."

[170] Actually, only 32 models were fit, because there were no item respondents in the 2002 survey for one of the levels of MULTRACECAT.

[171] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

For all logistic regression models, the starting list of covariates was the same as the list for the item response propensity models. Covariates were dropped in many cases to avoid convergence problems and instability problems. The sample size was very small for some of the models, because some of the levels of MULTRACECAT were relatively rare, especially the ones that were combinations of three or more major race categories. Pooling across the three survey years was done to increase the sample size. Due to an extremely small sample size and an unbalanced response variable for MULTRACECAT = 10 (black/African American, American Indian/AK Native, and Asian), no model was fit. Instead, the weighted percentage of respondents with each QD06RACE category was recorded for use in the next step. The final set of covariates for each logistic regression model is shown in Table E.4.

E.2.5 Final Assignment of Single Race

For the 2005 NSDUH respondents with MULTRACECAT levels 1 through 9 and 11, the parameter estimates from the previous step were used to estimate the probability that the respondent would have selected one of the component races as their "best race," if QD06 was in the questionnaire. For example, consider a 2005 NSDUH respondent with MULTRACECAT = 1 (white and black/African American only). Given the values of all covariates for this respondent and the parameter estimates from the model for MULTRACECAT = 1 from Step 4 (Fitting of Logistic Regression Models), the probability that this respondent would have chosen "white" had he or she been offered QD06 was estimated. The probability that this respondent would have chosen "black/African American" was simply one minus the probability that he or she would have chosen white. If this probability was x, then this respondent was assigned a single race of "white" with probability x, and "black/African American" with a probability of $1 - x$. The assignment was completed by comparing a randomly generated number to this probability. Taking the respondent with MULTRACECAT = 1 (white and black/African American only) as an example, if the uniform randomly generated number y [0,1] was less or equal to x, then this respondent would be given a single race of "white." Or, if y was greater than x, this respondent would be assigned to "black/African American."

For example, respondents from the 2005 survey with MULTRACECAT level 10 (black/African American, American Indian/AK Native, and Asian) were assigned "black/African American" with probability 0.385, "American Indian/AK Native" with probability 0.057, and "Asian" with probability 0.558. These numbers are the simple weighted proportions of each QD06RACE value using the pooled data from the 2000-2002 surveys.

The number of multiple-race respondents in the 2005 survey and their final single race assignments are summarized in Table E.5.

Table E.2 Results of Race Editing of Multiple-Race Respondents from the 2000-2002 Surveys

Survey Year	MULTRACECAT	QD06RACE	Number of Respondents
2000	White and Black/African American	White	48
		Black/African American	105
		Missing	125
	White and American Indian/AK Native	White	312
		American Indian/AK Native	106
		Missing	88
	White and Asian	White	137
		Asian	133
		Missing	33
	Black/African American and American Indian/AK Native	Black/African American	46
		American Indian/AK Native	8
		Missing	0
	Black/African American and Asian	Black/African American	21
		Asian	3
		Missing	2
	American Indian/AK Native and Asian	American Indian/AK Native	2
		Asian	5
		Missing	0
	White, Black/African American, and American Indian/AK Native	White	6
		Black/African American	15
		American Indian/AK Native	6
		Missing	9
	White, Black/African American, and Asian	White	0
		Black/African American	1
		Asian	1
		Missing	0
	White, American Indian/AK Native, and Asian	White	4
		American Indian/AK Native	0
		Asian	9
		Missing	1
	Black/African American, American Indian/AK Native, and Asian	Black/African American	2
		American Indian/AK Native	0
		Asian	2
		Missing	0
	White, Black/African American, American Indian/AK Native, and Asian	White	2
		Black/African American	3
		American Indian/AK Native	1
		Asian	2
		Missing	0

Table E.2 Results of Race Editing of Multiple-Race Respondents from the 2000-2002 Surveys (continued)

Survey Year	MULTRACECAT	QD06RACE	Number of Respondents
2001	White and Black/African American	White	33
		Black/African American	97
		Missing	235
	White and American Indian/AK Native	White	282
		American Indian/AK Native	115
		Missing	145
	White and Asian	White	138
		Asian	117
		Missing	65
	Black/African American and American Indian/AK Native	Black/African American	43
		American Indian/AK Native	8
		Missing	5
	Black/African American and Asian	Black/African American	14
		Asian	1
		Missing	8
	American Indian/AK Native and Asian	American Indian/AK Native	9
		Asian	7
		Missing	1
	White, Black/African American, and American Indian/AK Native	White	6
		Black/African American	16
		American Indian/AK Native	2
		Missing	8
	White, Black/African American, and Asian	White	2
		Black/African American	1
		Asian	2
		Missing	2
	White, American Indian/AK Native, and Asian	White	5
		American Indian/AK Native	2
		Asian	7
		Missing	2
	Black/African American, American Indian/AK Native, and Asian	Black/African American	5
		American Indian/AK Native	0
		Asian	0
		Missing	0
	White, Black/African American, American Indian/AK Native, and Asian	White	0
		Black/African American	2
		American Indian/AK Native	2
		Asian	2
		Missing	0

Table E.2 Results of Race Editing of Multiple-Race Respondents from the 2000-2002 Surveys (continued)

Survey Year	MULTRACECAT	QD06RACE	Number of Respondents
2002	White and Black/African American	White	51
		Black/African American	155
		Missing	156
	White and American Indian/AK Native	White	449
		American Indian/AK Native	180
		Missing	127
	White and Asian	White	178
		Asian	171
		Missing	41
	Black/African American and American Indian/AK Native	Black/African American	92
		American Indian/AK Native	20
		Missing	5
	Black/African American and Asian	Black/African American	26
		Asian	12
		Missing	7
	American Indian/AK Native and Asian	American Indian/AK Native	6
		Asian	8
		Missing	2
	White, Black/African American, and American Indian/AK Native	White	12
		Black/African American	29
		American Indian/AK Native	10
		Missing	10
	White, Black/African American, and Asian	White	0
		Black/African American	2
		Asian	3
		Missing	1
	White, American Indian/AK Native, and Asian	White	11
		American Indian/AK Native	2
		Asian	16
		Missing	0
	Black/African American, American Indian/AK Native, and Asian	Black/African American	2
		American Indian/AK Native	1
		Asian	2
		Missing	0
	White, Black/African American, American Indian/AK Native, and Asian	White	0
		Black/African American	0
		American Indian/AK Native	0
		Asian	0
		Missing	1

Table E.3 Summaries of Item Response Propensity Models

Survey Year	MULTRACECAT	Covariates
2000	White and Black/African American	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households
	White and American Indian/AK Native	Percentage Asian Population; Percentage Black/African American Population
	White and Asian	Percentage Asian Population
	Black/African American and American Indian/AK Native	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	Black/African American and Asian	Percentage Asian Population; Percentage Black/African American Population
	American Indian/AK Native and Asian	Age; Percentage of Owner-Occupied Households
	White, Black/African American, and American Indian/AK Native	Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White, Black/African American, and Asian	Age
	White, American Indian/AK Native, and Asian	Percentage Asian Population; Percentage American Indian/AK Native Population
	Black/African American, American Indian/AK Native, and Asian	Age; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American
	White, Black/African American, American Indian/AK Native, and Asian	Age; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Black/African American Population
2001	White and Black/African American	Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White and American Indian/AK Native	Percentage American Indian/AK Native Population
	White and Asian	Percentage Asian Population
	Black/African American and American Indian/AK Native	Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	Black/African American and Asian	Percentage Asian Population; Percentage Black/African American Population
	American Indian/AK Native and Asian	Percentage Asian Population; Percentage American Indian/AK Native Population
	White, Black/African American, and American Indian/AK Native	Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White, Black/African American, and Asian	Census Region
	White, American Indian/AK Native, and Asian	Census Region

Table E.3 Summaries of Item Response Propensity Models (continued)

Survey Year	MULTRACECAT	Covariates
2001 (cont'd)	Black/African American, American Indian/AK Native, and Asian	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households
	White, Black/African American, American Indian/AK Native, and Asian	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Black/African American Population
2002	White and Black/African American	Percentage Black/African American Population
	White and American Indian/AK Native	Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White and Asian	Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	Black/African American and American Indian/AK Native	Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	Black/African American and Asian	Percentage Asian Population; Percentage Black/African American Population
	American Indian/AK Native and Asian	Percentage Asian Population; Percentage American Indian/AK Native Population
	White, Black/African American, and American Indian/AK Native	Percentage of Owner-Occupied Households; Percentage Black/African American Population
	White, Black/African American, and Asian	Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White, American Indian/AK Native, and Asian	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	Black/African American, American Indian/AK Native, and Asian	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage Asian Population; Percentage of Owner-Occupied Households; Percentage American Indian/AK Native Population; Percentage Black/African American Population
	White, Black/African American, American Indian/AK Native, and Asian	N/A (no item respondents)

Table E.4 Summaries of Logistic Regression Models

MULTRACECAT	Covariates
White and Black/African American	Age; Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Asian Population; Percentage American Indian/AK Native Population; Percentage Black/African American Population
White and American Indian/AK Native	Age; Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Asian Population; Percentage American Indian/AK Native Population; Percentage Black/African American Population
White and Asian	Age; Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Asian Population; Percentage American Indian/AK Native Population; Percentage Black/African American Population
Black/African American and American Indian/AK Native	Age; Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage of Owner-Occupied Households; Percentage Asian Population; Percentage American Indian/AK Native Population; Percentage Black/African American Population
Black/African American and Asian	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage Asian Population; Percentage Black/African American Population
American Indian/AK Native and Asian	Census Region; Percentage of Owner-Occupied Households; Percentage Asian Population
White, Black/African American, and American Indian/AK Native	Census Region; Indicator of Whether the Householder Was Non-Hispanic/Latino Black/African American; Percentage American Indian/AK Native Population; Percentage Black/African American Population
White, Black/African American, and Asian	Percentage of Owner-Occupied Households
White, American Indian/AK Native, and Asian	Percentage of Owner-Occupied Households
Black/African American, American Indian/AK Native, and Asian	N/A
White, Black/African American, American Indian/AK Native, and Asian	Age

Table E.5 Summary of Multiple-Race Respondents and Their Final Single Race Assignments for the 2005 Survey

EDRACE	Final Single Race	Number of Respondents
White and Black/African American	White	96
	Black/African American	263
White and American Indian/AK Native	White	708
	American Indian/AK Native	296
White and Asian	White	202
	Asian	181
Black/African American and American Indian/AK Native	Black/African American	101
	American Indian/AK Native	30
Black/African American and Asian	Black/African American	31
	Asian	8
American Indian/AK Native and Asian	American Indian/AK Native	9
	Asian	9
White, Black/African American, and American Indian/AK Native	White	12
	Black/African American	50
	American Indian/AK Native	21
White, Black/African American, and Asian	White	5
	Black/African American	0
	Asian	9
White, American Indian/AK Native, and Asian	White	15
	American Indian/AK Native	1
	Asian	7
Black/African American, American Indian/AK Native, and Asian	Black/African American	3
	American Indian/AK Native	0
	Asian	2
White, Black/African American, American Indian/AK Native, and Asian	White	0
	Black/African American	2
	American Indian/AK Native	1
	Asian	1

Appendix F: Model Summaries

Appendix F: Model Summaries

F.1 Introduction

The tables in this appendix list the covariates used in all the imputation models that were run in the 2005 National Survey on Drug Use and Health (NSDUH).[172] 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 F.2 and the models for the drug variables are presented in Section F.3.With the exception of the lifetime usage models, separate tables are provided in Section F.3 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 given in Section F.4. Section F.5 presents the models for the income variables and Section F.6 presents the models for the health insurance variables. Also Section F.6 presents the models for both the "Old Method" and the "Constituent Variables Method," used to create the final imputation-revised health insurance variables. Chapter 10 provides a more detailed description of these two methods.

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 and not multiplication. In addition, when the abbreviation "MSA" is used, it represents "metropolitan statistical area."

F.1.1 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 detail in the main body of this report. No such descriptions are available for screener and segment-level variables, however. The following screener and segment-level variables were often used as covariates in both types of models for the PMN procedures.

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 household to one of three levels:

[172] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

Hispanic/Latino, non-Hispanic/Latino black/African American, or non-Hispanic/Latino non-black/African American.

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.

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 was adjusted to more recent data from Claritas, Inc.[173] This variable had five levels: segment in metropolitan statistical area (MSA) with 1 million or more persons; segment in MSA with 250,000 to 999,999 persons; segment in MSA with fewer than 250,000 persons; segment not in MSA and not in rural area; and segment not in MSA and in rural area.

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 71 percent, and more than 71 percent.

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 50 percent, and 50 percent or more.

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 also had three levels: less than 10 percent, 10 to 40 percent, and 40 percent or more.

Percentage Asian in Segment

The "percentage Asian 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 also had three levels: less than 5 percent, 5 to 10 percent, and 10 percent or more.

[173] Claritas, Inc., is a market research firm headquartered in San Diego, California.

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 also had three levels: less than 1 percent, 1 to 3 percent, and 3 percent or more.

F.2 Demographic Variables

For justifications of the aggregation of age groups for certain imputation steps, see Chapters 4 and 5.

Table F.1 Summaries for Response Propensity Models

Imputation Step	Variables Included in Response Propensity Model
Marital Status	Census Region; Gender; Population Density; Age Category; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Age Category * Gender
Race 12-17	Census Region; Household Type; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment
Race 18-25	Census Region; Household Type; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment
Race 26+	Census Region; Household Type; Age Category; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment
Hispanic/Latino Origin 12-17	Census Region; Imputation-Revised Race; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Hispanic/Latino Origin 18-25	Census Region; Imputation-Revised Race; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Hispanic/Latino Origin 26+	Census Region; Imputation-Revised Race; Age Category; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Hispanic/Latino Group	Census Region; Imputation-Revised Race; Gender; Age Category; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Age Category * Gender
Education Level 12-17	Census Region; Imputation-Revised Race; Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Education Level 18+	Census Region; Imputation-Revised Race; Gender; Age Category; Age Category * Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Employment Status 12-25	Census Region; Imputation-Revised Race; Gender; Age Category; Age Category * Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment

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

Imputation Step	Variables Included in Response Propensity Model
Employment Status 26+	Census Region; Imputation-Revised Race; Gender; Age Category; Age Category * Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Born in US 12-17	Gender; Imputation-Revised Race/Ethnicity; Percentage Owner Occupied in Segment; MSA; Census Region
Born in US 18-25	Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region
Born in US 26+	Gender; Age Category; Age Category * Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region
Age of Entry	Gender; Age Category; Age Category * Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Race/Ethnicity * Gender; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region

Table F.2 Summaries for Predictive Mean Models

Imputation Step	Variables Included in Predictive Mean Model
Marital Status	Age; Percentage Black/African American in Segment; Percentage Owner Occupied in Segment; Gender; Age * Gender; Census Region; Population Density; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment
Race 12-17	Census Region; Age; Age Squared; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment
Race 18-25	Census Region; Age; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Imputation-Revised Marital Status
Race 26+	Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Census Region
Hispanic/Latino Origin 12-17	Census Region; Imputation-Revised Race; Household Type; Age; Age Squared; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Hispanic/Latino Origin 18-25	Census Region; Imputation-Revised Race; Household Type; Age; Age Squared; Age Cubed; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Marital Status
Hispanic/Latino Origin 26+	Household Type; Census Region; Imputation-Revised Race; Age; Age Squared; Age Cubed; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Marital Status
Hispanic/Latino Group	Census Region; Imputation-Revised Race; Percentage Hispanic/Latino in Segment; Gender; Age; Age Squared; Age Cubed; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Owner Occupied in Segment; Age * Gender; Age Squared * Gender

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

Imputation Step	Variables Included in Predictive Mean Model
Education Level 12-17	Imputation-Revised Race; Gender; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Owner Occupied in Segment
Education Level 18+	Census Region; Imputation-Revised Race; Gender; Age; Age Squared; Age Cubed; Age * Gender; Age Squared * Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Marital Status
Employment Status 12-25	Census Region; Imputation-Revised Race; Gender; Age; Age Squared; Age * Gender; Age Squared * Gender; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Employment Status 26+	Age; Gender; Age * Gender; Imputation-Revised Marital Status; Imputation-Revised Race; Census Region; Percentage Black/African American in Segment; Percentage American Indian/AK Native in Segment; Percentage Asian in Segment; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Born in US 12-17	Gender; Age; Age Squared; Age * Gender; Age Squared * Gender; Imputation-Revised Race/Ethnicity; Percentage Owner Occupied in Segment; MSA; Census Region
Born in US 18-25	Gender; Age; Age Squared; Age * Gender; Age Squared * Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region
Born in US 26+	Gender; Age; Age Squared; Age * Gender; Age Squared * Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region
Age of Entry	Gender; Age; Age Squared; Age * Gender; Age Squared * Gender; Imputation-Revised Race/Ethnicity; Imputation-Revised Race/Ethnicity * Gender; Imputation-Revised Education Level; Imputation-Revised Employment Status; Imputation-Revised Marital Status; Percentage Owner Occupied in Segment; MSA; Census Region

F.3 Drug Variables

Table F.3 Lifetime Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Gender; Race; Gender * Race; MSA; Census Region; Cigarette Lifetime Indicator
18 to 25	Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; MSA; Census Region; Cigarette Lifetime Indicator
26+	Age Category; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; MSA; Census Region; Cigarette Lifetime Indicator

Table F.4 Cigarettes: 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Not applicable (N/A)	N/A
Recency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Daily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank Nondaily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank
Age at First Use	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Cigarettes 30-Day Frequency; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

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

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Race; Gender; Census Region; MSA	Intermediate Alcohol 30-Day Frequency; Imputation-Revised Age at First Use for Cigarettes; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.5 Cigarettes: 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Daily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank Nondaily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Cigarettes 30-Day Frequency; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

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

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Race; Gender; Census Region; MSA	Intermediate Alcohol 30-Day Frequency; Imputation-Revised Age at First Use for Cigarettes; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Marital Status; Education Level; Employment Status

Table F.6 Cigarettes: 26 Years and Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Daily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank Nondaily users: Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Cigarettes 30-Day Frequency; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.6 Cigarettes: 26 Years and Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Age Category; Race; Gender; Census Region; MSA	Intermediate Alcohol 30-Day Frequency; Imputation-Revised Age at First Use for Cigarettes; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Marijuana; Imputation-Revised Lifetime Indicator for Alcohol; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Marital Status; Education Level; Employment Status

Table F.7 Smokeless Tobacco (Chewing Tobacco and Snuff): 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A

Table F.7 Smokeless Tobacco (Chewing Tobacco and Snuff): 12- to 17-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Imputation-Revised Lifetime Indicators for Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Race; Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Chewing Tobacco daily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Census Region; Age; Race; MSA Chewing Tobacco nondaily users: Age; Race; Census Region; Imputation-Revised Lifetime Indicators for Alcohol and Hallucinogens Snuff daily users: Imputation-Revised Lifetime Indicators for Cigars and Stimulants; Age Snuff nondaily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Race; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Snuff 30-Day Frequency; Chewing Tobacco 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes and Daily Cigarettes; Imputation-Revised Recencies for Cigarettes, Snuff, and Chewing Tobacco; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.8 Smokeless Tobacco (Chewing Tobacco and Snuff): 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Age; Age Squared; Age Cubed; Race; MSA
12-Month Frequency	N/A	N/A

Table F.8 Smokeless Tobacco (Chewing Tobacco and Snuff): 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Chewing Tobacco daily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank Chewing Tobacco nondaily users: Age; Gender; Race; Age * Race; Census Region; Imputation-Revised Lifetime Indicator for Hallucinogens Snuff daily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Age Squared; Age Cubed; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank; Race; Gender * Race Snuff nondaily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank

Table F.8 Smokeless Tobacco (Chewing Tobacco and Snuff): 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Snuff 30-Day Frequency; Chewing Tobacco 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes and Daily Cigarettes; Imputation-Revised Recencies for Cigarettes, Snuff, and Chewing Tobacco; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.9 Smokeless Tobacco (Chewing Tobacco and Snuff): 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Smokeless Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; Education Level; Employment Status; Census Region; MSA; State Rank Chewing Tobacco: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; Age * Gender; Education Level; Census Region; State Rank Snuff: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Hallucinogens, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; Education Level; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A

Table F.9 Smokeless Tobacco (Chewing Tobacco and Snuff): 26-Year-Olds or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Imputation-Revised Lifetime Indicators for Pain Relievers and Cocaine; Race; Census Region; MSA Snuff: Imputation-Revised Lifetime Indicators for Cigars, Pipes, and Alcohol; Age Category; Race; Gender	Chewing Tobacco daily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Stimulants, Sedatives, and Heroin; Age; Race; Age * Race; Marital Status; Education Level; Employment Status; MSA Chewing Tobacco nondaily users: Age; Race; Age * Race; Imputation-Revised Recency for Cigarettes Snuff daily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank Snuff nondaily users: Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Snuff 30-Day Frequency; Chewing Tobacco 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes and Daily Cigarettes; Imputation-Revised Recencies for Cigarettes, Snuff, and Chewing Tobacco; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.10 Cigars: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff and Chewing Tobacco; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Cigars 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, and Smokeless Tobacco; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.11 Cigars: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff and Chewing Tobacco; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Cigars 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, and Smokeless Tobacco; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.12 Cigars: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff and Chewing Tobacco; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Heroin; Age Category; Race; Gender; Census Region	Imputation-Revised Recencies for Cigarettes and Smokeless Tobacco; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Cigars 30-Day Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, and Smokeless Tobacco; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.13 Pipes: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, and Cigars; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	No model used: no nonrespondents	No model used: no nonrespondents
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Table F.14 Pipes: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, and Cigars; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Alcohol, Marijuana, and Hallucinogens; Gender; Education Level	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Table F.15 Pipes: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, and Cigars; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Tranquilizers, and Stimulants; Gender; MSA	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Table F.16 Alcohol: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, and Pipes; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Alcohol 30-Day Frequency; Alcohol 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.17 Alcohol: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, and Pipes; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank

Table F.17 Alcohol: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Alcohol 30-Day Frequency; Alcohol 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.18 Alcohol: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, and Pipes; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Past Month Alcohol Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank

Table F.18 Alcohol: 26-Year-Olds or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Alcohol 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Alcohol 30-Day Frequency; Alcohol 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, and Cigars; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.19 Inhalants: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Inhalants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Inhalants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	Imputation-Revised Recencies for Cigarettes and Cigars; Race; Gender; Census Region; MSA	Intermediate Inhalants 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank

Table F.19 Inhalants: 12- to 17-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Inhalants 30-Day Frequency; Inhalants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, and Alcohol; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.20 Inhalants: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Heroin; Gender; Race; Marital Status; Education Level; Employment Status; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Gender; Age * Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Race; Gender * Race
12-Month Frequency	Intermediate Past Month Inhalants Indicator; Imputation-Revised Recencies for Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Inhalants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	Imputation-Revised Lifetime Indicators for Marijuana, Pain Relievers, Tranquilizers, and Sedatives	Age; Race; Education Level; Census Region; Intermediate Inhalants 12-Month Frequency

Table F.20 Inhalants: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Inhalants 30-Day Frequency; Inhalants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, and Alcohol; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.21 Inhalants: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Education Level; Age; Gender; Race; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Age; Gender; Education Level; Employment Status; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicators for Marijuana and Stimulants; Census Region	Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Age; Age Squared
12-Month Frequency	Intermediate Past Month Inhalants Indicator; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Gender	Age; Race; Education Level; Employment Status; Imputation-Revised Lifetime Indicators for Marijuana and Pain Relievers; Intermediate Past Month Inhalants Indicator
30-Day Frequency	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicator for Tranquilizers	Age; Employment Status; Imputation-Revised Lifetime Indicators for Pain Relievers and Crack; Intermediate Inhalants 12-Month Frequency
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Inhalants 30-Day Frequency; Inhalants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, and Alcohol; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.22 Marijuana: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank

Table F.22 Marijuana: 12- to 17-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Marijuana 30-Day Frequency; Marijuana 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, and Inhalants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.23 Marijuana: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank

Table F.23 Marijuana: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Marijuana 30-Day Frequency; Marijuana 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, and Inhalants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.24 Marijuana: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Past Month Marijuana Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Marijuana 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank

Table F.24 Marijuana: 26-Year-Olds or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Marijuana 30-Day Frequency; Marijuana 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, and Inhalants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.25 Hallucinogens: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Hallucinogens Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Hallucinogens Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Hallucinogens 12-Month Frequency; Imputation-Revised Recencies for Smokeless Tobacco, Cigars, Pipes, and Inhalants; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region	Intermediate Hallucinogens 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; MSA; State Rank

Table F.25 Hallucinogens: 12- to 17-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Hallucinogens 30-Day Frequency; Hallucinogens 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, and Marijuana; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.26 Hallucinogens: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Hallucinogens Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Hallucinogens Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank

Table F.26 Hallucinogens: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Intermediate Hallucinogens 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Hallucinogens 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Hallucinogens 30-Day Frequency; Hallucinogens 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, and Marijuana; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.27 Hallucinogens: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Education Level; Age; Gender; Race; Age Squared; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicator for Pain Relievers; Age; Age Squared; Age Cubed; Employment Status
Recency: past month vs. past year not past month	Race; MSA	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Inhalants, Pain Relievers, Stimulants, and Cocaine; Age; Age Squared; Race; Age * Race; Education Level; Employment Status; Census Region; MSA
12-Month Frequency	Intermediate Past Month Hallucinogens Indicator; Imputation-Revised Recencies for Smokeless Tobacco and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Census Region	Race; Education Level; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Tranquilizers and Stimulants
30-Day Frequency	Intermediate Hallucinogens 12-Month Frequency; Imputation-Revised Recencies for Smokeless Tobacco, Cigars, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, and Heroin	Age; Race; Marital Status; Intermediate Hallucinogens 12-Month Frequency

Table F.27 Hallucinogens: 26-Year-Olds or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Hallucinogens 30-Day Frequency; Hallucinogens 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, and Marijuana; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.28 Pain Relievers: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Pain Relievers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.29 Pain Relievers: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Pain Relievers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.30 Pain Relievers: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Past Month Pain Relievers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Pain Relievers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, and Hallucinogens; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.31 Tranquilizers: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Pain Relievers; Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Tranquilizers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.32 Tranquilizers: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A

Table F.32 Tranquilizers: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Tranquilizers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.33 Tranquilizers: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Stimulants, and Heroin; Gender; Age Category; MSA; State Rank	Imputation-Revised Lifetime Indicators for Hallucinogens, Stimulants, and Sedatives; Age; Marital Status; Education Level
12-Month Frequency	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; MSA	Intermediate Past Month Tranquilizers Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Tranquilizers 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.34 Stimulants: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Intermediate Lifetime Indicators for Inhalants and Hallucinogens; Age; Gender; Race; Age Squared; Gender * Race; Age * Gender; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Census Region; MSA	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Stimulants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Stimulants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Gender; Census Region; MSA	Stimulants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.35 Stimulants: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Stimulants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Stimulants Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A

Table F.35 Stimulants: 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Stimulants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.36 Stimulants: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Age Category; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicators for Pipes, Sedatives, and Heroin; Gender; Race; Gender * Race; Education Level; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA
12-Month Frequency	Imputation-Revised Recency for Pipes	Age; Age Squared; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicator for Heroin; Intermediate Past Month Stimulants Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Cigars, Pipes, Alcohol, Marijuana, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Age Category; Race; Gender; Census Region; MSA	Stimulants 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Sedatives, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.37 Sedatives: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Intermediate Lifetime Indicators for Inhalants and Hallucinogens; Age; Gender
Recency: past year vs. not past year	Imputation-Revised Recencies for Alcohol and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Gender; MSA	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, and Crack; Age; Age Squared; Age Cubed; Gender; Race; Age * Gender; Age * Race; Census Region; MSA
12-Month Frequency	Intermediate Past Month Sedatives Indicator; Imputation-Revised Recencies for Smokeless Tobacco, Cigars, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Gender	Intermediate Past Month Sedatives Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Pipes and Stimulants; Imputation-Revised Lifetime Indicator for Heroin; MSA	Sedatives 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.38 Sedatives: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Stimulants, and Heroin; Employment Status; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicator for Smokeless Tobacco; Age; Age Squared; Gender; Race; Marital Status; Education Level; Employment Status; Census Region
Recency: past month vs. past year not past month	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco and Cigars; Gender; Race; Marital Status; Education Level; Employment Status; Census Region	Imputation-Revised Recencies for Alcohol and Marijuana; Imputation-Revised Lifetime Indicator for Smokeless Tobacco; Age; Age Squared; Marital Status; Education Level; Employment Status; Census Region
12-Month Frequency	Intermediate Past Month Sedatives Indicator; Imputation-Revised Recencies for Cigarettes, Alcohol, Marijuana, and Pain Relievers; Imputation-Revised Lifetime Indicators for Crack and Heroin; Race	Intermediate Past Month Sedatives Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; State Rank
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Race; Gender; Census Region; MSA	Sedatives 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.39 Sedatives: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, and Crack; Gender; Race; Gender * Race; Marital Status; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Gender	Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicator for Cocaine; Age; Age Squared; Age Cubed; Gender; Marital Status
12-Month Frequency	Gender; Census Region	Gender; Marital Status; Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicator for Cocaine; Intermediate Past Month Sedatives Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Marijuana, Hallucinogens, Pain Relievers, and Sedatives; Imputation-Revised Lifetime Indicators for Cocaine and Crack; Age Category; Race; Gender; Census Region; MSA	Sedatives 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Cocaine, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.40 Cocaine (and Crack): 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cocaine: Intermediate Lifetime Indicator for Tranquilizers Crack: Intermediate Lifetime Indicator for Hallucinogens
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Gender; Race; Gender * Race; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Stimulants; Imputation-Revised Lifetime Indicators for Crack and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Cocaine 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Hallucinogens, and Stimulants; Imputation-Revised Lifetime Indicator for Crack; Race; Gender; Census Region; MSA	Age; Census Region; Imputation-Revised Recency for Cigars; Intermediate Cocaine 12-Month Frequency
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Crack and Heroin; Race; Gender; Census Region; MSA	Cocaine 30-Day Frequency; Cocaine 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA

Table F.41 Cocaine (and Crack): 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cocaine: Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region Crack: Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
12-Month Frequency	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Race; Gender; Census Region; MSA	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank

Table F.41 Cocaine (and Crack): 18- to 25-Year-Olds (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicator for Heroin; Race; Gender; Census Region; MSA	Intermediate Cocaine 12-Month Frequency; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Census Region; Age; Gender; Race; Age Squared; Age Cubed; Age * Race; Gender * Race; Age * Gender; Marital Status; Education Level; Employment Status; MSA; State Rank
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Crack and Heroin; Race; Gender; Census Region; MSA	Cocaine 30-Day Frequency; Cocaine 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.42 Cocaine (and Crack): 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	<u>Cocaine:</u> Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region <u>Crack:</u> Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Education Level; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; State Rank; MSA; Census Region
Recency	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Gender; Age Category; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin; Gender; Race; Gender * Race; Marital Status; Education Level; Employment Status; Census Region; MSA
12-Month Frequency	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicator for Heroin; Age Category; Race; Gender; Census Region; MSA	Intermediate Past Month Cocaine Indicator; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
30-Day Frequency	Intermediate Cocaine 12-Month Frequency; Imputation-Revised Recencies for Cigars, Pipes, and Hallucinogens; Age Category; Race	Age; Gender; Race; Gender * Race; Imputation-Revised Lifetime Indicator for Crack; Intermediate Cocaine 12-Month Frequency

Table F.42 Cocaine (and Crack): 26-Year-Olds or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Crack and Heroin; Age Category; Race; Gender; Census Region; MSA	Cocaine 30-Day Frequency; Cocaine 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cocaine; Imputation-Revised Lifetime Indicators for Daily Cigarettes, Crack, and Heroin; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

Table F.43 Heroin: 12- to 17-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Chewing Tobacco, Cigars, and Alcohol; Age
Recency: past year vs. not past year	Imputation-Revised Lifetime Indicators for Pipes and Sedatives; Race	Imputation-Revised Lifetime Indicators for Pipes, Sedatives, and Crack; Census Region
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicator for Sedatives; Race	Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicator for Crack; Age; Age Squared; Age Cubed
12-Month Frequency	Intermediate Past Month Heroin Indicator; Imputation-Revised Recencies for Cigarettes, Hallucinogens, and Pain Relievers	Imputation-Revised Recency for Marijuana; Intermediate Past Month Heroin Indicator
30-Day Frequency	Intermediate Heroin 12-Month Frequency; Imputation-Revised Recencies for Cigarettes and Alcohol	Age; Imputation-Revised Recencies for Marijuana, Sedatives, and Crack
Age at First Use	Imputation-Revised Recencies for Cigarettes, Alcohol, Hallucinogens, and Heroin; Gender	Heroin 30-Day Frequency; Heroin 12-Month Frequency; Imputation-Revised Age at First Use for Cigarettes; Imputation-Revised Recencies for Cigarettes, Alcohol, Hallucinogens, and Heroin; Gender; Census Region; MSA

Table F.44 Heroin: 18- to 25-Year-Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Age; Gender; Race; Age Squared; Age Cubed; Marital Status; Employment Status; Gender * Race; Age * Gender; Age * Race; MSA; Census Region
Recency: past year vs. not past year	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Gender; Race; Education Level; Employment Status; Census Region; MSA; State Rank	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Age; Age Squared; Age Cubed; Gender; Race; Age * Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Sedatives, and Crack; Census Region; MSA	Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Cocaine, and Crack; Age; Age Squared; Gender; Race; Age * Gender; Marital Status; Education Level; Employment Status
12-Month Frequency	Intermediate Past Month Heroin Indicator; Imputation-Revised Recencies for Tranquilizers, Stimulants, and Crack; Census Region	Imputation-Revised Recency for Marijuana; Intermediate Past Month Heroin Indicator
30-Day Frequency	Intermediate Heroin 12-Month Frequency; Imputation-Revised Recencies for Tranquilizers and Cocaine	Race; MSA; Imputation-Revised Recencies for Tranquilizers and Crack; Intermediate Heroin 12-Month Frequency
Age at First Use	Imputation-Revised Recencies for Hallucinogens and Tranquilizers; Race	Heroin 30-Day Frequency; Heroin 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes and Cocaine; Imputation-Revised Recencies for Cigarettes, Pipes, Stimulants, and Cocaine; Imputation-Revised Lifetime Indicator for Daily Cigarettes; Age; Race; State Rank; Age Squared; Age * Race; Education Level; Employment Status

Table F.45 Heroin: 26-Year-Olds or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Table F.3*	Intermediate Lifetime Indicators for Cocaine and Crack; Age; Gender
Recency: past year vs. not past year	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Crack; Gender; Employment Status; State Rank	Imputation-Revised Recencies for Alcohol and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Sedatives, Cocaine, and Crack; Employment Status; Census Region; MSA
Recency: past month vs. past year not past month	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Hallucinogens, Tranquilizers, and Stimulants	Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, and Sedatives
12-Month Frequency	Intermediate Past Month Heroin Indicator; Imputation-Revised Recencies for Alcohol and Stimulants; Age Category	Race; Marital Status; Imputation-Revised Recencies for Marijuana and Cocaine; Intermediate Past Month Heroin Indicator
30-Day Frequency	Intermediate Heroin 12-Month Frequency; INTERCEP	Gender; Employment Status; Imputation-Revised Recency for Cocaine; Intermediate Heroin 12-Month Frequency
Age at First Use	Imputation-Revised Recencies for Cigarettes, Cigars, and Marijuana; Age Category; Race; Census Region	Heroin 30-Day Frequency; Heroin 12-Month Frequency; Imputation-Revised Ages at First Use for Cigarettes, Daily Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Lifetime Indicator for Daily Cigarettes; Age; Gender; Race; State Rank; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Marital Status; Education Level; Employment Status; Census Region; MSA

F.4 Household Composition Variables

Table F.46 Household Composition: 12- to 17-Year-Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Gender; Race; Gender * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Total People in Household (Screener)	Age; Total People in Household (Screener); Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size	Age; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Number of Persons Older Than 64 Years Old in Household (HH65)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment	Gender; Race; Gender * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment
Number of Respondent's Family Members in Household	Age; Age Squared; Gender; Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment

Table F.46 Household Composition: 12- to 17-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Number of Respondent's Family Members in Household Younger Than 18 Years Old	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment

Table F.47 Household Composition: 18- to 25-Year-Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Education Level; Employment Status; Total People in Household (Screener)	Age; Total People in Household (Screener); Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size	Age; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Older Than 64 Years Old in Household (HH65)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Gender; MSA; Percentage Owner Occupied in Segment; Education Level; Employment Status

Table F.47 Household Composition: 18- to 25-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Number of Respondent's Family Members in Household	Age; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Respondent's Family Members in Household Younger Than 18 Years Old	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status

Table F.48 Household Composition: 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Total People in Household (Screener)	Age; Total People in Household (Screener); Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size	Age; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Older Than 64 Years Old in Household (HH65)	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age Squared; Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Education Level; Employment Status

Table F.48 Household Composition: 26 to 64 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Number of Respondent's Family Members in Household	Race; Percentage Hispanic/Latino in Segment	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Respondent's Family Members in Household Younger Than 18 Years Old	Gender; Race; Gender * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status

Table F.49 Household Composition: 65+ Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level	Age; Total People in Household (Screener); Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Gender; Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size	Age; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Older Than 64 Years Old in Household (HH65)	Gender; Race; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other Family Present in Household (FAMSKIP)	Gender; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status	Age; Gender; Race; Census Region; Marital Status; Employment Status
Number of Respondent's Family Members in Household	Gender; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status

Table F.49 Household Composition: 65+ Year Olds (continued)

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Number of Respondent's Family Members in Household Younger Than 18 Years Old	Gender; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level	Gender; Race; Gender * Race; Age * Race; Census Region; Imputation-Revised Number of Respondent's Family Members in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Family in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status

F.5 Income Variables

Table F.50 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12 to 17	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Age Cubed * Gender; Age Cubed * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank
18 to 25	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Age Cubed * Gender; Age Cubed * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; 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; Income State Rank
26 to 64	Gender; Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; 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; Income State Rank
65+	Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Income State Rank

Table F.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12- to 17-Year-Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank
Supplemental Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security
Wages	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support
Food Stamps	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income
Welfare Payments	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security
Welfare Services	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments

Table F.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12- to 17-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
# Welfare Months	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps
Investment Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services
Child Support	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income
Other Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages

Table F.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12- to 17-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps

Table F.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18- to 25-Year-Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education Level; Employment Status
Supplemental Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security
Wages	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support
Food Stamps	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income
Welfare Payments	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security

Table F.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18- to 25-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Services	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments
# Welfare Months	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status
Investment Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services
Child Support	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income

Table F.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18- to 25-Year-Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Other Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages
Total Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status

Table F.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status
Supplemental Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security
Wages	Age; Gender; Race; Age Squared; Gender * Race; Age * Gender; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support
Food Stamps	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income
Welfare Payments	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security
Welfare Services	Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education Level; Intermediate Family Welfare Payments

Table F.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
# Welfare Months	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status
Investment Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services
Child Support	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income
Other Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages

Table F.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status

Table F.54 Dichotomous Income Indicators in Predictive Mean Modeling: 65+ Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Income State Rank; Education Level
Supplemental Security	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security
Wages	Gender; Race; Census Region; MSA; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Intermediate Family Child Support
Food Stamps	Age; Race; Age Squared; Age Cubed; Census Region; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income
Welfare Payments	Age; Gender; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Income State Rank; Education Level
Welfare Services	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments
# Welfare Months	Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status

Table F.54 Dichotomous Income Indicators in Predictive Mean Modeling: 65+ Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Investment Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services
Child Support	Age; Gender; Race; Census Region; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Income State Rank; Marital Status; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Services; Intermediate Family Investment Income
Other Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages
Total Income	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wages; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education Level; Employment Status

Table F.55 Income Finer Categories in Response Propensity Models

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
12 to 17	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
26 to 64	Gender; Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
65+	Gender; Race; Gender * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)

Table F.56 Income Finer Categories in Predictive Mean Models

Age Group	Variables Included in Income Models (Finer Categorization)
12 to 17	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status
26 to 64	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status
65+	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age * Race; Age Squared * Gender; Age Squared * Race; Census Region; MSA; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Percentage Owner Occupied in Segment; 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; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status

F.6 Health Insurance Variables

Table F.57 Health Insurance, Constituent Variables Method: Response Propensity Models

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
12 to 17	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode
	Other Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode
18 to 25	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; Marital Status; Education Level; Employment Status; MSA; Percentage Owner Occupied in Segment; Income Recode
	Other Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; Marital Status; Education Level; Employment Status; MSA; Percentage Owner Occupied in Segment; Income Recode
26 to 64	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; Marital Status; Education Level; Employment Status; MSA; Percentage Owner Occupied in Segment; Income Recode
	Other Health Insurance[1]	Gender; Race; Gender * Race; Education Level; Employment Status; MSA; Percentage Owner Occupied in Segment; Income Recode; Marital Status
65+	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Gender; Race; Marital Status; Education Level; MSA; Percentage Owner Occupied in Segment; Income Recode
	Other Health Insurance[1]	Gender; Race; Gender * Race; Education Level; Employment Status; MSA; Percentage Owner Occupied in Segment; Income Recode; Marital Status

[1]The 26 to 64 and 65 years or older age groups were included in the same response propensity model for Other Health Insurance.

Table F.58 Health Insurance, Constituent Variables Method: Predictive Mean Models, 12- to 17-Year-Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Imputation-Revised Household Size
Medicare	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Family Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode; Family Other Income; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Family Other Income; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Imputation-Revised Household Size

Table F.59 Health Insurance, Constituent Variables Method: Predictive Mean Models, 18- to 25-Year-Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Imputation-Revised Household Size
Medicare[1]	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Family Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Personal Other Income; Lifetime Military Service; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Family Other Income; Other Family Members in Household; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Imputation-Revised Household Size

[1]The 18 to 25 and 26 to 64 age groups were included in the same predictive mean model for Medicare.

Table F.60 Health Insurance, Constituent Variables Method: Predictive Mean Models, 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Employment Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Imputation-Revised Household Size
Medicare[1]	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Family Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Personal Other Income; Lifetime Military Service; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Family Other Income; Other Family Members in Household; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance[2]	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Imputation-Revised Household Size

[1]The 18 to 25 and 26 to 64 age groups were included in the same predictive mean model for Medicare.
[2]The 26 to 64 and 65 years or older age groups were included in the same predictive mean model for other health insurance.

Table F.61 Health Insurance, Constituent Variables Method: Predictive Mean Models, 65+ Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender; Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Household Size
Medicare	Age; Age Squared; Gender; Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Personal Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age Squared; Gender; Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Personal Other Income; Lifetime Military Service; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age Squared; Gender; Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Family Other Income; Other Family Members in Household; Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance[1]	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Percentage Owner Occupied in Segment; Income Recode; Family Wages; Family Participation in Government Assistance Programs; Family Social Security; Family Investment Income; Other Family Members in Household; Imputation-Revised Household Size

[1]The 26 to 64 and 65 years or older age groups were included in the same predictive mean model for other health insurance.

Table F.62 Old Method Health Insurance, Based on INSUR3: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
18 to 25	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
26 to 64	Age; Gender; Race; Gender * Race; Age * Gender; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
65+	Gender; Race; Gender * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size

Table F.63 Old Method Health Insurance, Based on INSUR: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
18 to 25	Age; Age Squared; Age Cubed; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
26 to 64	Age; Age Squared; Gender; Race; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
65+	Gender; Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Imputation-Revised Household Size

Table F.64 Old Method Health Insurance: Predictive Mean Models, 12- to 17-Year-Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3.

Table F.65 Old Method Health Insurance: Predictive Mean Models, 18- to 25-Year-Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3.

Table F.66 Old Method Health Insurance: Predictive Mean Models, 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3.

Table F.67 Old Method Health Insurance: Predictive Mean Models, 65+ Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Gender * Race; Age * Gender; Age * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Marital Status; Education Level; Employment Status; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age Squared; Age Cubed; Gender * Race; Age * Gender; Age Squared * Gender; Age * Race; Age Squared * Race; MSA; Percentage Owner Occupied in Segment; Percentage Hispanic/Latino in Segment; Percentage Non-Hispanic/Latino Black/African American in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3.

Appendix G: Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

Appendix G: Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

G.1 Introduction

For all the 2005 National Survey on Drug Use and Health (NSDUH)[174] variables for which imputations were implemented using predictive mean neighborhoods (PMN), whether the method was univariate (UPMN) or multivariate (MPMN), restrictions were placed upon the neighborhood prior to the assignment of imputed values. The pool of potential donors for a given recipient was restricted so that donors and recipients were as alike as possible (likeness constraints), and so that the donor's values were consistent with the preexisting nonmissing values of the recipient (logical constraints). Logical constraints (summarized in Appendix H) were not loosened, because this would have resulted in an inconsistency that would not have been countenanced.[175] However, some likeness constraints were loosened, even though this resulted in donors and recipients being less alike in various cases. If no donors were available under the most stringent set of constraints, the likeness constraints were loosened, one at a time, until a donor was found. This appendix summarizes the number of cases for which donors were available under each of the various likeness constraints, starting with the most stringent constraint. The appendix is organized by groups of variables requiring imputation using the PMN method: demographics, lifetime use of drugs, recency and frequency of drug use, age at first drug use, household roster, income, and health insurance. Each table demonstrates the number of recipients who found corresponding donors that satisfied the set of likeness constraints utilized in each try. Rows in a table contain the likeness constraints applied in each try, which starts with the most stringent set of constraints. Likeness constraints are usually systematically removed one at a time as the tries progress. The "Frequency" in the table is the number of recipients for whom donors were found in a given try. The labels for some of the likeness constraints given in the tables are not self-evident; therefore, more complete descriptions are given in the following paragraphs.

Although statistical imputation of the drug use or income variables could not have proceeded separately within each State due to insufficient pools of donors, information about the State of residence of each respondent was incorporated in the PMN procedure. For the drug use variables, in the hot-deck step of PMN, respondents were separated into three State 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 into another, and the remainder into a third category. The States were separated into three income groups for the income variables, depending upon the proportion of families 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 into another category, and the remainder into 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

[174] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[175] Logical constraints define what is normally referred to as an "imputation class."

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.

The phrase "Donor's predicted means each within *x* percent of recipient's predicted means" appears in each of the tables corresponding to a multivariate imputation, and the phrase "Donor's predicted mean 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. It also defines the neighborhood. Once this constraint was loosened, the neighborhood was abandoned and the candidate with the predicted mean closest to the recipient's (subject to the constraints that were still on the pool of donors) was chosen as the donor.

G.2 Demographics

Tables G.1 through G.5 present information on the likeness constraints applied during the imputation procedures for the core demographic variables: marital status, race, Hispanic origin, Hispanic group, and education level. Tables G.6 through G.8 present information on the likeness constraints for the noncore demographic variables: employment status, indicator of birth in the United States, and immigrant age of entry into the United States. The segment-level variable, SEGID (Segment ID) that was used only in the likeness constraints for demographic imputation is described here.

Segment ID

As described in the 2005 NSDUH sample design report (Morton, Chromy, Hunter, & Martin, 2006) 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, due to the large number of levels, it was used as a constraint in the hot-deck step of the PMN procedure for race, Hispanicity, education, and employment status as noted in Chapters 4 and 5 of this report. For more information regarding segments, see the 2005 NSDUH sample design report (Morton et al., 2006).

G.2.1 Marital Status Variables

Table G.1 Marital Status Imputations

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) Donor's age within 3 years of recipient's age (B) Donor's predicted means within 5 percent of recipient's predicted means	3	5	2
(A) Donor's age within 3 years of recipient's age	0	0	2

G.2.2 Race Variables

Table G.2 Race Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means (C) If the recipient was Hispanic/Latino nonspecific, the donor must have been of Hispanic/Latino origin[1] (D) If the recipient selected one or more Hispanic/Latino categories: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spaniard, the donor's Hispanic/Latino group value must equal to one of the Hispanic/Latino groups mentioned by the recipient[1] (E) If the recipient was non-Hispanic/Latino Mexican, the donor must be Mexican (but the donor could have been Hispanic/Latino or non-Hispanic/Latino)[1]	176	156	167
(A) Donor's predicted means within 5 percent of recipient's predicted means (B) If the recipient was Hispanic/Latino nonspecific, the donor must have been of Hispanic/Latino origin (C) If the recipient selected one or more Hispanic/Latino categories: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spaniard, the donor's Hispanic/Latino group value must equal to one of the Hispanic/Latino groups mentioned by the recipient (D) If the recipient was non-Hispanic/Latino Mexican, the donor must be Mexican (but the donor could have been Hispanic/Latino or non-Hispanic/Latino)	326	312	285
(A) If the recipient was Hispanic/Latino nonspecific, the donor must have been of Hispanic/Latino origin (B) If the recipient selected one or more Hispanic/Latino categories: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spaniard, the donor's Hispanic/Latino group value must equal to one of the Hispanic/Latino groups mentioned by the recipient (C) If the recipient was non-Hispanic/Latino Mexican, the donor must be Mexican (but the donor could have been Hispanic/Latino or non-Hispanic/Latino)	108	152	4

[1] These likeness constraints are never loosened.

G.2.3 Hispanic Origin Variables

Table G.3 Hispanic Origin Imputations

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) Segment of donor = Segment of recipient (B) Donor's predicted mean within 5 percent of recipient's predicted mean	101	7	0
(A) Donor's predicted mean within 5 percent of recipient's predicted mean	37	7	4

G.2.4 Hispanic/Latino Group Variables

Table G.4 Hispanic/Latino Group Imputations

Likeness Constraints	Frequency[1]
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	15
(A) Donor's predicted means within 5 percent of recipient's predicted means	35
None	3

[1] The hot-deck program for Hispanic/Latino Group is not separated into age groups.

G.2.5 Education Variables

Table G.5 Education Imputations

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	1	1	1
(A) Donor's predicted means within 5 percent of recipient's predicted means	3	0	7
None	0	0	1

G.2.6 Employment Variables

Table G.6 Employment Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	0	3	1
(A) Donor's predicted means within 5 percent of recipient's predicted means	6	17	14
None	0	0	2

G.2.7 Immigrant Variables

Table G.7 Indicator of Birth in the United States Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	2	1	1
(A) Donor's predicted means within 5 percent of recipient's predicted means	7	4	8

Table G.8 Age of Entry into the United States Imputation

Likeness Constraints	Frequency[1]
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means (C) Donor's age of entry less than recipient's current age[2] (D) The difference between recipient's current age and the donor's age of entry ≤1 if the recipient lived in United States less than a year; or the difference >1 if the recipient lived in United States more than a year[2]	0
(A) Donor's predicted means within 5 percent of recipient's predicted means (B) Donor's age of entry less than recipient's current age (C) The difference between recipient's current age and the donor's age of entry ≤1 if the recipient lived in United States less than a year; or the difference >1 if the recipient lived in United States more than a year	28

[1] The hot-deck program for immigrant age of entry is not separated into age groups.

[2] These likeness constraints are never loosened.

G.3 Drug Variables

The imputation of the drug use variables was done separately for three age groups: 12 to 17, 18 to 25, and 26 or older. For each of the drugs, a multivariate imputation was done for the recency and frequency variables, and a univariate imputation was done for the age at first use variable(s). The tables in this appendix show the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

G.3.1 Likeness Constraints for Lifetime Imputation

Table G.9 Lifetime Imputations

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Lifetime use of donor = Lifetime use of recipient for each nonmissing lifetime indicator	457	98	63
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) 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	73	64	40
(A) State rank of donor = State rank of recipient (B) 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	37	23	25

[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 methamphetamines and other stimulants. The cocaine family includes cocaine as a whole and crack (although it is impossible to be missing the lifetime indicator for cocaine as a whole, but have a nonmissing response for crack).

G.3.2 Likeness Constraints for Recency and Frequency Imputation, by Drug

Tables G.10 through G.23 present information on the likeness constraints for recency and frequency imputation for the following drugs: tobacco (i.e., cigarettes, smokeless tobacco [chewing tobacco and snuff], cigars, and pipes), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, and sedatives), cocaine, and heroin.

Table G.10 Cigarette Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	347	128	19
(A) State rank of donor = State rank of recipient	20	8	4

Table G.11 Smokeless Tobacco Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's recencies for chewing tobacco and snuff agree with recipient's recencies (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means	111	102	13
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's lifetime use statuses for chewing tobacco and snuff agree with recipient's lifetime use statuses (when nonmissing)[1]	12	4	0
(A) State rank of donor = State rank of recipient (B) Donor's lifetime use statuses for chewing tobacco and snuff agree with recipient's lifetime use statuses (when nonmissing)[1]	45	28	11

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included here for clarity.

Table G.12 Cigar Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	209	149	39
(A) State rank of donor = State rank of recipient	19	5	4

Table G.13 Pipe Recency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted probability of past month use within 5 percent of recipient's predicted probability of past month use	0	1	3

Table G.14 Alcohol Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	536	607	384
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	357	159	119

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.15 Inhalants Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	84	8	2
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	241	45	8

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.16 Marijuana Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	104	87	38
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	186	127	44

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.17 Hallucinogens Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's recencies for LSD, PCP, and Ecstasy agree with recipient's recencies (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means (D) Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens (E) Donor's recency must match recipient's recency (when nonmissing)[1]	3	9	2
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens (D) Donor's recency must match recipient's recency (when nonmissing)[1]	5	10	0
(A) State rank of donor = State rank of recipient (B) Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens (C) Donor's recency must match recipient's recency (when nonmissing)[1]	113	108	42
(A) Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens (B) Donor's recency must match recipient's recency (when nonmissing)[1]	0	0	1

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.18 Pain Relievers Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's recency for OxyContin agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means (D) Donor's recency must match recipient's recency (when nonmissing)[1]	106	76	23
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	3	6	1
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	175	113	58

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.19 Tranquilizers Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	11	18	8
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	51	55	27

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.20 Stimulants Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's recency for methamphetamines agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means (D) Donor's recency must match recipient's recency (when nonmissing)[1]	8	10	0
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	3	0	0
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	75	45	16

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.21 Sedatives Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	1	1	1
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	29	8	7

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.22 Cocaine Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's recency for crack agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means (D) Donor's recency must match recipient's recency (when nonmissing)[1]	3	26	8
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	1	17	1
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	62	91	52
(A) Donor's recency must match recipient's recency (when nonmissing)[1]	0	3	0

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

Table G.23 Heroin Recency and Frequency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means (C) Donor's recency must match recipient's recency (when nonmissing)[1]	0	3	1
(A) State rank of donor = State rank of recipient (B) Donor's recency must match recipient's recency (when nonmissing)[1]	5	14	3

[1] Although this constraint is also used as a logical constraint for some missingness patterns, it is included for clarity.

G.3.3 Likeness Constraints for Age-at-First-Use Imputation, by Drug

Tables G.24 through G.38 present information on the likeness constraints for age at first use (AFU) imputation for the following drugs: tobacco (i.e., cigarettes, cigarette daily use, smokeless tobacco [chewing tobacco and snuff], and cigars), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, and sedatives), cocaine, and heroin. Table G.25 presents information for the imputation of the ever-daily-used variable for cigarettes. Although this is not an age–at-first-use variable, it is summarized here because its imputation procedures are completed immediately before cigarette age at first daily use.

Table G.24 Cigarette Age-at-First-Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	299	122	87
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	0	5
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same	0	0	3
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then recipient was same; if recipient was not a past 3 years user, then recipient was same[1]	0	0	2

[1] Although this is a logical constraint, it is included for clarity.

Table G.25 Cigarette Ever-Daily-Used Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	7	6	4
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	2	2

Table G.26 Cigarette Age at First Daily Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	15	21	68
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	0	9
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past three years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same	3	0	6
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then recipient was same; if recipient was not a past 3 years user, then recipient was same[1]	0	0	1
(A) AFU of donor ≤ Age of recipient,[1] Age of donor ≥ Age of recipient (B) If recipient was not a past year user, then recipient was same; if recipient was not a past 3 years user, then recipient was same[1]	1	0	2

[1] Although this is a logical constraint, it is included for clarity.

Table G.27 Smokeless Tobacco Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	88	106	43
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	17	6	11
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same	36	8	14
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then recipient was same; if recipient was not a past-3-years user, then recipient was same[1]	2	0	4
(A) AFU of donor ≤ Age of recipient,[1] Age of donor ≥ Age of recipient (B) If recipient was not a past year user, then recipient was same; if recipient was not a past-3-years user, then recipient was same[1]	5	0	1

[1] Although this is a logical constraint, it is included for clarity.

Table G.28 Cigar Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	164	155	168
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	3	1	17
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was a past 3 years not past year user, then donor was the same; if recipient was lifetime not past 3 years user, then donor was the same	0	0	13

Table G.29 Alcohol Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	341	154	203
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	14
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	0	0	5

Table G.30 Inhalants Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	305	63	30
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	2	1
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	1	1	2

Table G.31 Marijuana Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	103	52	56
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	2
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	0	0	1

Table G.32 Hallucinogens Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy) (D) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (E) Donor's predicted mean within 5 percent of recipient's predicted mean	52	52	46
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy) (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (D) Donor's predicted mean within 5 percent of recipient's predicted mean	9	5	12
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy)[2] (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1] (D) Donor's predicted mean within 5 percent of recipient's predicted mean	6	3	11
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy)[2] (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1]	2	3	8
(A) Age of donor = Age of recipient (B) If recipient was a lifetime not past year user, then donor was not a past year user (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy)[2] (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1]	1	3	5

Table G.32 Hallucinogens Age at First Use Imputation (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens),[1] Donor was at least as old as recipient, but no more than 20 years older than recipient (B) If recipient was a lifetime not past year user, then donor was also not a past year user (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy)[2] (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1]	1	0	3
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens),[1] Donor was no more than 20 years older than recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for overall hallucinogens, LSD, PCP, and Ecstasy)[2] (C) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1]	1	0	2
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens),[1] Donor was no more than 20 years older than recipient (B) Donor agrees with recipient with respect to lifetime use for LSD, PCP, and Ecstasy (checked only if recipient is a nonrespondent for LSD, PCP, and/or Ecstasy AFU, as applicable)[1]	1	0	0

[1] Although this is a logical constraint, it is included for clarity.

[2] These constraints were intended to match recency of use for parent and child drugs. They were not applied exactly as intended, because lifetime nonusers of child drugs 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 G.33 Pain Relievers Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall pain relievers and OxyContin) (D) Donor agrees with recipient with respect to lifetime use for OxyContin (E) Donor's predicted mean within 5 percent of recipient's predicted mean	257	219	128
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall pain relievers and OxyContin) (C) Donor agrees with recipient with respect to lifetime use for OxyContin (D) Donor's predicted mean within 5 percent of recipient's predicted mean	3	3	20
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall pain relievers and OxyContin)[2] (C) Donor agrees with recipient with respect to lifetime use for OxyContin (checked only if recipient is a nonrespondent for OxyContin AFU)[1] (D) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	3
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall pain relievers and OxyContin)[2] (C) Donor agrees with recipient with respect to lifetime use for OxyContin (checked only if recipient is a nonrespondent for OxyContin AFU)[1]	2	1	18
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for both overall pain relievers and OxyContin)[2] (C) Donor agrees with recipient with respect to lifetime use for OxyContin (checked only if recipient is a nonrespondent for OxyContin AFU)	0	0	2

Table G.33 Pain Relievers Age at First Use Imputation (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) AFU of donor ≤ Age of recipient (for overall pain relievers),[1] Donor was at least as old as recipient, but no more than 20 years older than recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for overall pain relievers and OxyContin)[2] (C) Donor agrees with recipient with respect to lifetime use for OxyContin (checked only if recipient is a nonrespondent for OxyContin AFU)[1]	0	0	1

[1] Although this is a logical constraint, it is included for clarity.

[2] These constraints were intended to match recency of use for parent and child drugs. They were not applied exactly as intended, because lifetime nonusers of child drugs 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 G.34 Tranquilizers Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	48	55	45
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	5	0	13
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	0	1	2
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then recipient was same[1]	0	0	1
(A) AFU of donor ≤ Age of recipient,[1] Age of donor ≥ Age of recipient (B) If recipient was not a past year user, then recipient was same[1]	0	0	2

[1] Although this is a logical constraint, it is included for clarity.

Table G.35 Stimulants Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall stimulants and Methamphetamines) (D) Donor agrees with recipient with respect to lifetime use for Methamphetamines (E) Donor's predicted mean within 5 percent of recipient's predicted mean	47	48	30
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall stimulants and Methamphetamines) (C) Donor agrees with recipient with respect to lifetime use for Methamphetamines (D) Donor's predicted mean within 5 percent of recipient's predicted mean	14	4	10
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall stimulants and Methamphetamines)[1] (C) Donor agrees with recipient with respect to lifetime use for Methamphetamines (checked only if recipient is a nonrespondent for Methamphetamines AFU)[2] (D) Donor's predicted mean within 5 percent of recipient's predicted mean	1	1	3
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall stimulants and Methamphetamines)[1] (C) Donor agrees with recipient with respect to lifetime use for Methamphetamines (checked only if recipient is a nonrespondent for Methamphetamines AFU)[2]	5	1	9
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for both overall stimulants and Methamphetamines)[1] (C) Donor agrees with recipient with respect to lifetime use for Methamphetamines (checked only if recipient is a nonrespondent for Methamphetamines AFU)	0	0	1

Table G.35 Stimulants Age at First Use Imputation (continued)

	Frequency		
Likeness Constraints	**12-17**	**18-25**	**26+**
(A) AFU of donor ≤ Age of recipient (for overall stimulants),[2] Donor was at least as old as recipient, but no more than 20 years older than recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for overall stimulants and methamphetamines)[1] (C) Donor agrees with recipient with respect to lifetime use for methamphetamines (checked only if recipient is a nonrespondent for methamphetamines AFU)[2]	0	0	2

[1] These constraints were intended to match recency of use for parent and child drugs. They were not applied exactly as intended, since lifetime nonusers of child drugs were not correctly considered. However, the practical impact was small since child drug nonusers could not be donors for a missing child drug age at first use.

[2] Although this is a logical constraint, it is included for clarity.

Table G.36 Sedatives Age at First Use Imputation

	Frequency		
Likeness Constraints	**12-17**	**18-25**	**26+**
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	20	13	13
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	5	1	7
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	2	0	4
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then recipient was same[1]	0	0	0
(A) AFU of donor ≤ Age of recipient,[1] Donor was at least as old as recipient (B) If recipient was not a past year user, then recipient was same[1]	0	0	1
(A) AFU of donor ≤ Age of recipient,[1] (B) If recipient was not a past year user, then recipient was same[1]	0	0	1

[1] Although this is a logical constraint, it is included for clarity.

Table G.37 Cocaine Age at First Use Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall cocaine and crack) (D) Donor agrees with recipient with respect to lifetime use for crack (E) Donor's predicted mean within 5 percent of recipient's predicted mean	15	34	39
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall cocaine and crack) (C) Donor agrees with recipient with respect to lifetime use for crack (D) Donor's predicted mean within 5 percent of recipient's predicted mean	2	0	2
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall cocaine and crack)[1] (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)[2] (D) Donor's predicted mean within 5 percent of recipient's predicted mean	1	0	0
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime not past year user, then donor was either a lifetime not past year user or a lifetime nonuser (this check is done for both overall cocaine and crack)[1] (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)[2]	0	0	0
(A) Age of donor = Age of recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for both overall cocaine and crack)[1] (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)	0	0	1
(A) AFU of donor ≤ Age of recipient (for overall cocaine),[2] Donor was at least as old as recipient, but no more than 20 years older than recipient (B) If recipient was not a past year user, then donor was not a past year user (this check is done for overall cocaine and crack)[1] (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)[2]	1	0	0

[1] These constraints were intended to match recency of use for parent and child drugs. They were not applied exactly as intended, because lifetime nonusers of child drugs 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.

[2] Although this is a logical constraint, it is included for clarity.

Table G.38 Heroin Age at First Use Imputation

Likeness Constraints	Frequency		
	12-17	**18-25**	**26+**
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (D) Donor's predicted mean within 5 percent of recipient's predicted mean	1	3	1
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same (C) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	0
(A) Age of donor = Age of recipient (B) If recipient was a past year user, then donor was a past year user; if recipient was not a past year user, then donor was same	0	0	1

G.4 Household Composition (Roster) Variables

Tables G.39 through G.44 present information on the likeness constraints applied during the imputation procedures for the six household composition (roster) variables: IRHHSIZE, IRKID17, IRHH65, IRFAMSKP, IRFAMSZE, and IRKIDFAM.

Table G.39 IRHHSIZE Imputations

Likeness Constraints	Frequency			
	12-17	**18-25**	**26-64**	**65+**
Donor's predicted mean within 5 percent of recipient's predicted mean	14	14	17	2
None	1	0	0	1

Table G.40 IRKID17 Imputations

Likeness Constraints	Frequency			
	12-17	**18-25**	**26-64**	**65+**
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	69	101	61	5
(A) IRHHSIZE of donor = IRHHSIZE of recipient	3	1	0	2
None	0	0	1	0

Table G.41 IRHH65 Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	224	165	60	5
(A) IRHHSIZE of donor = IRHHSIZE of recipient	1	0	0	1
None	0	0	1	1

Table G.42 IRFAMSKP Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRKID17 of donor = IRKID17 of recipient (C) If recipient was married, then donor was married; otherwise if the recipient was not currently married, then donor was not currently married[1]	16	16	19	1
(A) IRKID17 of donor = IRKID17 of recipient (B) If recipient was married, then donor was married; otherwise if the recipient was not currently married, then donor was not currently married[1]	1	1	0	0

[1] This constraint is a likeness constraint that is never loosened.

Table G.43 IRFAMSZE Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient (C) IRKID17 of donor = IRKID17 of recipient	28	22	28	2
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	0	1	0	1
(A) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	0	0
None	1	0	1	0

Table G.44 IRKIDFAM Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient (C) IRKID17 of donor = IRKID17 of recipient (D) IRFAMSZE of donor = IRFAMSZE of recipient	36	23	20	1
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient (C) IRFAMSZE of donor = IRFAMSZE of recipient	1	0	0	0
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRFAMSZE of donor = IRFAMSZE of recipient	0	0	0	0
(A) IRFAMSZE of donor = IRFAMSZE of recipient	3	4	2	0

G.5 Income Variables

G.5.1 Binary Variable Phase

Six 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 six variables (i.e., welfare payments, welfare services, food stamps, binary income, investment 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. In the tables, these six variables are referred to as "welfare-correlated variables." All of the other likeness constraints that were applied are self-explanatory in the tables.

Table G.45 Binary Income Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Age of donor = Age of recipient (B) Donor's values for welfare-correlated edited binary income variables are the same as recipient's values (when nonmissing or logically assigned) (C) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (D) Donor's predicted means within 5 percent of recipient's predicted means for all missing family variables (E) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing or logically assigned), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing or logically assigned) (F) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	1250	1512	552	82
(A) Age of donor = Age of recipient (B) Donor's values for welfare-correlated edited binary income variables are the same as recipient's values (when nonmissing or logically assigned) (C) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (D) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing or logically assigned), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing or logically assigned) (E) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	534	699	416	135

Table G.45 Binary Income Imputations (continued)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Age of donor is within 5 years of age of recipient (B) Donor's values for welfare-correlated edited binary income variables are the same as recipient's values (when nonmissing or logically assigned) (C) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (D) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing or logically assigned), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing or logically assigned) (E) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	11	23	37	13
(A) Donor's values for welfare-correlated edited binary income variables are the same as recipient's values (when nonmissing or logically assigned) (B) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (C) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing or logically assigned), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing or logically assigned) (D) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	0	0	8	1

Table G.45 Binary Income Imputations (continued)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (B) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing or logically assigned), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing or logically assigned) (C) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	1	5	6	2
(A) If recipient is missing other-family edited variable of a (personal, other family) pair, both the donor's value and the recipient's value for personal edited variable have to be 2 if nonmissing (B) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to family welfare payments (if nonmissing) and family welfare services (if nonmissing or logically assigned) (C) If recipient is missing any level (personal or other family) edited variables of PAYMENT, SERVICE, CHILD SUPPORT, or WAGES or is missing other-family edited variable of SOCIAL SECURITY, then the donor must match the recipient with respect to those highly correlated covariates, such as whether there are adults aged 65 years or older, kids younger than 18, or adults aged 18-64 in the household and employment status	0	0	0	4

G.5.2 Specific Category Phase

Table G.46 Specific Income Imputations

Likeness Constraints	Frequency			
	12-17	**18-25**	**26-64**	**65+**
(A) Donor's predicted mean within 10 percent of recipient's predicted mean (B) PINC2 of donor = PINC2 of recipient, if nonmissing (C) FINC2 of donor = FINC2 of recipient, if nonmissing (D) IRPINC1 of donor = IRPINC1 of recipient (E) IRFINC1 of donor = IRFINC1 of recipient	2700	3016	1776	418
(A) PINC2 of donor = PINC2 of recipient, if nonmissing (B) FINC2 of donor = FINC2 of recipient, if nonmissing (C) IRPINC1 of donor = IRPINC1 of recipient (D) IRFINC1 of donor = IRFINC1 of recipient	3	3	1	7

G.6 Health Insurance Variables

Table G.47 presents information on the likeness constraints for the health insurance variables created using the "Old Method." The remaining tables present information for the health insurance variables created using the "Constituent Variables Method." See Chapter 10 for an explanation of the two methods. Briefly, in the Constituent Variables Method, four variables (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT) were imputed simultaneously in an MPMN program, and one variable (IROTHHLT) was imputed in a UPMN program, following the imputation of other four variables. 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. As a result, there is at least one table for each missingness pattern. The final table in this section (Table G.71) presents the likeness constraints applied in the UPMN program for IROTHHLT.

In several instances in these health insurance tables, variable names are used without description for the purposes of brevity. (See Chapter 10 for greater 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 4-level categorical family-income variable, with levels <$20K, $20K to <$50K,

$50K to <$75K, and $75K or over), IRFAMIN1 (a 2-level family income variable with levels <$20K and $20K or over), IRFAMOTH/IRPOTH (an indicator of whether the respondent's family in the household or the respondent himself/herself received income from sources other than those considered in the income questions of the questionnaire), and IRFAMSOC/IRPSOC (an indicator of whether the respondent's family in the household or the respondent himself/herself received income from social security). For the latter two sets of variables, the match between donors and recipient was attempted on the personal income variable if the respondent was 18 or older. However, if the respondent was under 18, the match was attempted on the family income variable.

Table G.47 Health Insurance (IRINSUR, IRINSUR3) and Private Health Insurance (IRPINSUR) Imputations, Old Method

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Age of donor = Age of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	384	185	35	4
(A) Age of donor = Age of recipient	6	0	3	2

Table G.48 Health Insurance, Constituent Variables Method (MPMN), Only CAIDCHIP Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for CAIDCHIP within 5 percent of recipient's predicted mean (B) GOVTPROG of donor = GOVTPROG of recipient (C) MEDICARE, CHAMPUS, and PRVHLTIN of donor = MEDICARE, CHAMPUS, and PRVHLTIN of recipient	138	60	18	7
(A) Donor's predicted mean for CAIDCHIP within 5 percent of recipient's predicted mean (B) GOVTPROG of donor = GOVTPROG of recipient	0	3	0	1

Table G.49 Health Insurance, Constituent Variables Method (MPMN), Only MEDICARE Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for MEDICARE within 5 percent of recipient's predicted mean (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) CAIDCHIP, CHAMPUS, and PRVHLTIN of donor = CAIDCHIP, CHAMPUS, and PRVHLTIN of recipient	24	28	8	1
(A) Donor's predicted mean for MEDICARE within 5 percent of recipient's predicted mean (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient	0	1	0	0

Table G.50 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and MEDICARE Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) CHAMPUS and PRVHLTIN of donor = CHAMPUS and PRVHLTIN of recipient	11	6	6	0
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient	1	4	3	0
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	3	0	0	0
None	1	0	1	0

Table G.51 Health Insurance, Constituent Variables Method (MPMN), Only CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for CHAMPUS within 5 percent of recipient's predicted mean (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) CAIDCHIP, MEDICARE, and PRVHLTIN of donor = CAIDCHIP, MEDICARE, and PRVHLTIN of recipient	30	20	6	0
(A) Donor's predicted mean for CHAMPUS within 5 percent of recipient's predicted mean (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	1	0	0	0

Table G.52 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) MEDICARE and PRVHLTIN of donor = MEDICARE and PRVHLTIN of recipient	14	4	1	0
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	3	0	0	0
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	2	0	0	0

Table G.53 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for MEDICARE and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (F) CAIDCHIP and PRVHLTIN of donor = CAIDCHIP and PRVHLTIN of recipient	1	0	0	0

Table G.54 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (G) PRVHLTIN of donor = PRVHLTIN of recipient	7	1	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	1	0	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	0	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	2	1	0	1

Table G.55 Health Insurance, Constituent Variables Method (MPMN), Only PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) INCOME of donor = INCOME of recipient (C) CAIDCHIP, MEDICARE, and CHAMPUS of donor = CAIDCHIP, MEDICARE, and CHAMPUS of recipient	117	63	10

Table G.56 Health Insurance, Constituent Variables Method (MPMN), Only PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) IRFAMIN1 of donor = IRFAMIN1 of recipient (C) CAIDCHIP, MEDICARE, and CHAMPUS of donor = CAIDCHIP, MEDICARE, and CHAMPUS of recipient	3

Table G.57 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) INCOME of donor = INCOME of recipient (D) MEDICARE and CHAMPUS of donor = MEDICARE and CHAMPUS of recipient	44	9	1
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) INCOME of donor = INCOME of recipient	0	1	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	1	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means	1	1	0
None	1	0	0

Table G.58 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient (D) MEDICARE and CHAMPUS of donor = MEDICARE and CHAMPUS of recipient	0

Table G.59 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Youngest Three Age Groups

<table>
<tr><th rowspan="2">Likeness Constraints</th><th colspan="3">Frequency</th></tr>
<tr><th>12-17</th><th>18-25</th><th>26-64</th></tr>
<tr><td>(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means
(B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then
a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability
b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14
(C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient
(D) INCOME of donor = INCOME of recipient
(E) CAIDCHIP and CHAMPUS of donor = CAIDCHIP and CHAMPUS of recipient</td><td>4</td><td>5</td><td>0</td></tr>
<tr><td>(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means
(B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then
a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability
b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14
(C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient
(D) INCOME of donor = INCOME of recipient</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means
(B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then
a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability
b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14
(C) IRFAMIN1 of donor = IRFAMIN1 of recipient</td><td>0</td><td>0</td><td>0</td></tr>
<tr><td>(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means
(B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then
a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability
b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14</td><td>0</td><td>0</td><td>0</td></tr>
</table>

Table G.59 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Youngest Three Age Groups (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26-64
If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	0	2	0

Table G.60 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient (D) CAIDCHIP and CHAMPUS of donor = CAIDCHIP and CHAMPUS of recipient	0

Table G.61 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) INCOME of donor = INCOME of recipient (F) CHAMPUS of donor = CHAMPUS of recipient	7	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	1	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	0	0	0
None	7	3	1

Table G.62 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRPSOC of donor = IRPSOC of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient (E) CHAMPUS of donor = CHAMPUS of recipient	0

Table G.63 Health Insurance, Constituent Variables Method (MPMN), CHAMPUS and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) INCOME of donor = INCOME of recipient (E) CAIDCHIP and MEDICARE of donor = CAIDCHIP and MEDICARE of recipient	12	4	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) INCOME of donor = INCOME of recipient	0	1	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	0	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	0	0

Table G.64 Health Insurance, Constituent Variables Method (MPMN), CHAMPUS and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRPOTH of donor = IRPOTH of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient (E) CAIDCHIP and MEDICARE of donor = CAIDCHIP and MEDICARE of recipient	0

Table G.65 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) INCOME of donor = INCOME of recipient (F) MEDICARE of donor = MEDICARE of recipient	22	3	1
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	1	0	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	16	4	0

Table G.66 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient (F) MEDICARE of donor = MEDICARE of recipient	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	1

Table G.67 Health Insurance, Constituent Variables Method (MPMN), MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (F) INCOME of donor = INCOME of recipient (G) CAIDCHIP of donor = CAIDCHIP of recipient	3	0	0

Table G.68 Health Insurance, Constituent Variables Method (MPMN), MEDICARE, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency 65+
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) IRPSOC of donor = IRPSOC of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient (F) CAIDCHIP of donor = CAIDCHIP of recipient	0

Table G.69 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (G) INCOME of donor = INCOME of recipient	18	0	2
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then a. If the recipient has no job due to disability (JBSTATR = 14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	25	14	9

Table G.70 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRPSOC of donor = IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRPOTH of donor = IRPOTH of recipient (F) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	1

Table G.71 Health Insurance, Constituent Variables Method (UPMN), Any Other Health Insurance

Likeness Constraints	Frequency		
	12-17	18-25	26+
Donor's predicted mean within 5 percent of recipient's predicted mean	74	109	15

Appendix H: Missingness Patterns

Appendix H: Missingness Patterns

H.1 Introduction

For the majority of variables that had missing values imputed in the 2005 National Survey on Drug Use and Health (NSDUH),[176] the imputation method used was predictive mean neighborhoods (PMN). Some of these variables were imputed in sets. Specifically, an item nonrespondent with missing values for more than one variable in the set received values for all missing variables from the same donor. This is referred to as a "multivariate assignment." On the other hand, some variables were imputed one at a time using a "univariate assignment." In addition, some of the variables were imputed using a predictive mean vector with more than one element (multivariate matching), while others were imputed using a predictive mean vector with only one element (univariate matching). For variables that were binary or continuous and were not part of a multivariate set, the predictive mean vector and the assignment of imputed values were both univariate. However, multinomial variables that were not part of a multivariate set were imputed using a multivariate vector of predicted means (from a multinomial logistic model), from which a single imputed value (the level of the categorical variable) was imputed. A multivariate set of variables could be imputed based on a single univariate model. This could occur if the variables were all inextricably related, whereby a model from one of the variables was sufficient to describe the responses for all the characteristics of interest. In most cases, a multivariate predictive mean vector was used to match donors and recipients for a multivariate set of response variables. Table H.1 provides examples of variables that were imputed using each of the four methods.

Table H.1 Examples of Variables Imputed Using Each of the Four Methods of PMN

	Variables Imputed One at a Time (Univariate Assignment)	Variables Imputed in a Set (Multivariate Assignment)
Predictive mean vector has one element (univariate matching)	IRHOIND, IRHHSIZE, IRHH65, IRKID17, IRFAMSKP, IRMJAGE	{IRPINC2, IRFINC2, IRFAMIN2}, {IRCOCAGE, IRCRKAGE}
Predictive mean vector has more than one element (multivariate matching)	IRMARIT, IRHOGRP4, EMPSTATY, IREDUC	{IRRACE2, IRNWRACE}, {lifetime drug use}, {IRHERRC, IRHERFY, IRHERFM}, {binary sources of income}, {IRINSUR, IRINSUR3, IRPINSUR}, {IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT}

For many of these variables, the item nonrespondents were segregated into missingness patterns, which were simply patterns of nonresponse. Missingness patterns arose in two ways. The first occurred for sets of variables that underwent multivariate assignment: item nonrespondents were segregated into missingness patterns based on which variables were

[176] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

missing. The second way occurred when logical editing restricted an item nonrespondent to only a subset of the variable's possible values. For example, logical editing sometimes restricted a lifetime user of a drug to past year use; in these cases, the recipient received a final imputed value of 1 or 2 for drug recency. This could happen for any variable(s) that underwent multivariate matching.

This appendix focuses on the variables, or sets of variables, for which the set of logical constraints and/or the predictive mean vector differed between missingness patterns. It is limited to variables to which the PMN method was applied. The other imputation methods used in the 2005 survey were not multivariate. The tables in this appendix specify the following for each missingness pattern:

1) the number of item nonrespondents exhibiting the pattern ("Number of Cases");

2) the set of logical constraints applied to the potential donors ("Logical Constraints"); and

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 constraints as described in Appendix F.

Often, differences between missingness patterns with respect to the predictive mean vector were due to the use of conditional probabilities. If something about the item nonrespondent was known, then probabilities conditioned on what was known were used. For example, only past month users were included in models for 30-day frequency. Therefore, the predictive means calculated using these models were conditional on past month use of the drug. If an item nonrespondent was missing both recency and 30-day frequency for that drug, probabilities conditional on lifetime use, not on past month use, were used for the predictive mean vector. Conditional probabilities often resulted if the variables, which were imputed using a multivariate assignment method, were related in a hierarchical manner, such as overall health insurance and private health insurance in the "Old Method" (see Chapter 10 for details). Also, these types of conditional probabilities occurred if partial information was available about an item nonrespondent, such as the cases where it was known that the recipient was a past year user of a drug, but it was unknown whether he or she was a past month user.

Section H.2 shows the variable or set of variables that used missingness patterns along with logical constraints and predictive mean vectors, as appropriate. Some tables also give the number of item nonrespondents showing each missingness pattern. Section H.2.1 deals with race; Section H.2.2, with employment status; Section H.2.3, with drug lifetime use; Section H.2.4, with drug recency and frequency; Section H.2.5, with the source of income variables; and Sections H.2.6, with the health insurance variables.

H.2 Tables Showing Missingness Patterns and the Restrictions on the Set of Potential Donors

A few items to note regarding the tables in Section H.2 are as follows. In the missingness pattern section of the tables, a blank entry in the columns indicates that all information was available. An entry of "Missing" indicates that all information was missing. Other entries in the missingness pattern section give the available information, indicating that the information was partially missing. However, if the entry is shown in parentheses, all information was present and additional details are shown in the respective table.

H.2.1 Race

Table H.2 illustrates missingness patterns for race imputation in the 2005 study.

Table H.2 Restrictions and Portion of the Predictive Mean Vector for Race

#	Missingness Pattern	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Completely missing	1,643	None	1. R1 2. R2 3. R3
2	Known to be Asian	6	Donor must be Asian in part or in full	1. R1 2. R2 3. R3
3	Known to be multiple race, but no other information	20	Donor must be more than one Race	1. R1 2. R2 3. R3
4	Known to be nonwhite, but no other information	12	Donor must not be white only	1. R1 2. R2 3. R3
5	Known to be white or both white and American Indian/Alaska Native	1	Donor must be white only, or white and American Indian/ Alaska Native only	1. R1/(R1+R3)
6	Known not to be American Indian/ Alaska Native, in part or in full	0	Donor must not be American Indian/Alaska Native, in part or in full	1. R1/(R1+R2+R4) 2. R2/(R1+R2+R4)
7	Known to be non-Hispanic Mexican	4	Donor must be Mexican (Hispanic or non-Hispanic)	1. R1 2. R2 3. R3

[1]The predictive mean vector components are defined by the following:
1. R1 = P(White)
2. R2 = P(Black)
3. R3 = P(American Indian/Alaska native)

H.2.2 Employment Status

Table H.3 illustrates the two missingness patterns for employment status.

Table H.3 Restrictions and Portion of the Predictive Mean Vector for Employment Status

#	Missingness Pattern	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Completely missing	19	None	1. E1 2. E2 3. E3
2	Known to be employed; part-time vs. full-time status unknown	24	Donor must be employed	1. E1/(E1+E2)

[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)

H.2.3 Drug Lifetime Use

There were a large number of missingness patterns for drug lifetime use. The questionnaire included 15 "gate questions" that corresponded to lifetime use variables, plus several "subgate" 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. Apart from these restrictions, any combination of the lifetime drug variables could be missing.

Only one logical constraint was utilized in the multivariate imputation of lifetime use. If item nonrespondents were known to have used pain relievers, but both their OxyContin and "other" pain reliever indicators were missing, they were required to have a donor who was a lifetime user of pain relievers. This pattern of nonresponse occurs when respondents respond affirmatively to PR04 but fail to select any drugs from the card in PR04a.

The probabilities associated with the 14 gate questions (Table H.4) 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.

Table H.4 Elements of Full Predictive Mean Vector for Drug Lifetime Use

Lifetime Drug Use	Predictive Mean
Heroin Lifetime	P(Lifetime User)
Crack Lifetime	P(Lifetime User)
Cocaine Lifetime	P(Lifetime User)
Sedatives Lifetime	P(Lifetime User)
Stimulants/Methamphetamines Lifetime	P(Lifetime User)
Tranquilizers Lifetime	P(Lifetime User)
Pain Relievers/OxyContin Lifetime	P(Lifetime User)
Hallucinogens/LSD/PCP/Ecstasy Lifetime	P(Lifetime User)
Marijuana Lifetime	P(Lifetime User)
Inhalants Lifetime	P(Lifetime User)
Alcohol Lifetime	P(Lifetime User)
Pipes Lifetime	P(Lifetime User)
Snuff/Chewing Tobacco Lifetime	P(Lifetime User)
Cigars Lifetime	P(Lifetime User)

H.2.4 Drug Recency and Frequency

Tables H.4 to H.21 on the following pages illustrate missingness patterns for drug recency and frequency of use. In this section, pain relievers, sedatives, and tranquilizers had identical missingness patterns and are therefore presented in the same table. 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."

Table H.5 Constraints for Tobacco (Cigarettes and Cigars)

Constraint #	Logical Constraint
Tob1	Donor must have used within the past 3 years (a recency category of 1, 2, or 3)
Tob2	Donor cannot be a past month user (recency cannot equal 1)
Tob3	Donor must have used drug within the past year (recency = 1 or 2)
Tob4	Donor must be a past month user (recency = 1)
Tob5	If the recipient was never a daily user of cigarettes (CIGDLYMO = 2), the donor's 30-day cigarette frequency cannot equal 30
Tob6	If recipient's age at first use equals his or her current age, the donor's 30-day frequency (1) cannot be greater than the number of days between the recipient's interview date and his or her date of first drug use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his or her birthday (inclusive)
Tob7	Donor must be a past year (but not past month) user, or a past 3 years (but not past year) user (recency = 2 or 3)

Table H.6 Restrictions and Portion of the Predictive Mean Vector for Cigarette Users

Missingness Pattern			Number of Cases	Logical Constraints	Predictive Mean Vector[2]
#	Recency	30-Day Frequency[1]			
1	Past year	Missing	14	(Tob3,5)	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)
2	Missing (lifetime use known)	Missing	18	(Tob5)	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM
2	Missing (lifetime use imputed)	Missing	0		
3	(Past month)	Missing	22	(Tob4-6)	1. D 2. PM
4	Not past year		213	(Tob3,5)	1. R3/(R3+R4)
5	Not past month		97	(Tob2,5)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
6	Past year but not past month, or past 3 years but not past year		157	(Tob5,7)	1. R2/(R2+R3)
7	Past 3 years	Missing	5	(Tob1,5)	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*D)/(R1+R2+R3) 4. R1*(1-D)*PM/(R1+R2+R3)
	30-day frequency logically assigned based on estimated value, no missing values.		142	(Tob1,5)	
	Lifetime user, nothing missing		38,023	(None)	
	Imputed to lifetime nonuse		0	(None)	
	Lifetime nonuser, nothing missing		29,617	(None)	

[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 use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. R3 = P(past 3 years but not past year use | lifetime use)
4. D = P(daily use | past month use)
5. PM = P(use on a given day in the past month | past month use)

Table H.7 Restrictions and Portion of the Predictive Mean Vector for Cigar Users

Missingness Pattern			Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Recency	30-Day Frequency			
1	Past year	Missing	20	(Tob3)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
2	Missing (Lifetime use known)	Missing	7	(none)	1. R1 2. R2 3. R3 4. R1*PM
2	Missing (Lifetime use imputed)	Missing	5		
3	(Past month)	Missing	10	(Tob4,6)	1. PM
4	Not past year		153	(Tob3)	1. R3/(R3+R4)
5	Not past month		85	(Tob2)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
6	Past year but not past month, or past 3 years but not past year		142	(Tob7)	1. R2/(R2+R3)
7	Past 3 years	Missing	3	(Tob1)	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*PM)/(R1+R2+R3)
	30-day frequency logically assigned based on estimated value, no missing values.		43		
	Lifetime user, nothing missing		21,831		
	Imputed to lifetime nonuse		9		
	Lifetime nonuser, nothing missing		46,000		

[1] The predictive mean vector components are defined by the following:

1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. R3 = P(past 3 years but not past year use | lifetime use)
4. PM = P(use on a given day in the past month | past month use)

Table H.8 Constraints for Smokeless Tobacco (Chewing Tobacco and Snuff)

Constraint #	Description
SLT1	Donor must have used chew within the past 3 years (a recency category of 1, 2, or 3)
SLT2	Donor must have used snuff within the past 3 years (a recency category of 1, 2, or 3)
SLT3	If donor is not a chew user, then recipient must also not be a chew user (and vice versa)
SLT4	If donor is not a snuff user, then recipient must also not be a snuff user (and vice versa)
SLT5	If recipient's age at first chew use equals his or her current age, the donor's 30-day chew frequency (1) cannot be greater than the number of days between the recipient's interview date and his or her date of first chew use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his or her birthday (inclusive)
SLT6	If recipient's age at first snuff use equals his or her current age, the donor's 30-day snuff frequency (1) cannot be greater than the number of days between the 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 the recipient's interview date and his or her birthday (inclusive)
SLT7	Donor must be a past month chew user (chew recency = 1)
SLT8	Donor must be a past month snuff user (snuff recency = 1)
SLT9	Donor's snuff recency must equal recipient's snuff recency
SLT10	Donor's chew recency must equal recipient's chew recency
SLT11	Donor must have used chew within the past year (snuff recency = 1 or 2)
SLT12	Donor must have used snuff within the past year (chew recency = 1 or 2)
SLT13	Donor must be a past 3 years (but not past year) or lifetime (but not past 3 years) chew user (chew recency = 3 or 4)
SLT14	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)
SLT15	Donor must be a past year (but not past month), past 3 years (but not past year), or lifetime (but not past 3 years) chew user (chew recency = 2, 3, or 4)
SLT16	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)
SLT17	Donor must be a past year (but not past month) or a past 3 years (but not past year) chew user (chew recency = 2 or 3)
SLT18	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 H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	(Past month)	(Past month)	Missing	Missing	0	(SLT3-8)	1. DC 2. PMC 3. DS 4. PMS
2	(Past month)		Missing		2	(SLT3-5,7,9)	1. DC 2. PMC
3		(Past month)		Missing[1]	3	(SLT3-4,6,8,10)	1. DS 2. PMS
4		Missing (Lifetime use known)		Missing	5	(SLT3-4,6,10)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
4		Missing (Lifetime use imputed)		Missing	4		
5	(Past month)	Missing (Lifetime use known)	Missing	Missing	0	(SLT3-6,10)	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
5	(Past month)	Missing (Lifetime use imputed)	Missing	Missing	0		
6	Missing (lifetime use known)		Missing		1	(SLT3-5,9)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
6	Missing (lifetime use imputed)		Missing		2		
7	Missing (lifetime use known)	(Past month)	Missing	Missing	0	(SLT3-6,8)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS
7	Missing (lifetime use imputed)	(Past month)	Missing	Missing	0		

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
8		Past year		Missing	9	(SLT3-4,10-11)	1. R1/(R1+R2) 2. RS1*DS/(RS1+RS2) 3. RS1*(1-DS)*PMS/(RS1+RS2)
9	Past year		Missing		4	(SLT3-5,8,12)	1. R1/(R1+R2) 2. RC1*DC/(RC1+RC2) 3. RC1*(1-DC)*PMC/(RC1+RC2)
10	Missing (lifetime use known)	Missing (Lifetime use known)	Missing	Missing	1	(SLT3-6)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
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		
11	Not past year				48	(SLT3-4,8,13)	1. R3/(R3+R4)
12		Not past year			73	(SLT3-4,10,14)	1. R3/(R3+R4)
13	Not past year	Not past year			10	(SLT3-4,13-14)	1. R3/(R3+R4)
14	Not past month				28	(SLT3-4,9,15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern					Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.			
15		Not past month			42	(SLT3-4,10,16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
16	Not past month	Not past month			2	(SLT3-4,15-16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
17	Not past month	(Past month)		Missing	0	(SLT3-4,6,8,15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DS 4. PMS
18	(Past month)	Not past month	Missing		0	(SLT3-5,7,16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC
19	Not past month	Missing (lifetime use known)		Missing	1	(SLT3-4,6,15)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
19	Not past month	Missing (lifetime use imputed)		Missing	0		
20	Missing (lifetime use known)	Not past month	Missing		0	(SLT3-5,16)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
20	Missing (lifetime use imputed)	Not past month	Missing		0		
21	Not past month	Not past year			0	(SLT3-4,14-15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
22	Not past year	Not past month			0	(SLT3-4,13,16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
23	(Lifetime use of snuff, chewing tobacco, or both missing in raw data. Missing values imputed to nonuse in lifetime imputation; nothing missing at this point in sequence)				0		

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern					Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.			
24	Not past year	Missing (lifetime use known)		Missing	0	(SLT3-4,6,13)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
24	Not past year	Missing (lifetime use imputed)		Missing	0		
25	Missing (lifetime use known)	Not past year	Missing		0	(SLT3-5,14)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
25	Missing (lifetime use imputed)	Not past year	Missing		0		
26	Past year	Past year	Missing	Missing	1	(SLT3-6,11-12)	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)
27		Past 3 years		Missing	2	(SLT2-4,10)	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RS1*DS/ (RS1+RS2+RS3) 4. RS1*(1-DS)*PMS/ (RS1+RS2+RS3)
28	Past 3 years		Missing		0	(SLT1,3-4,9)	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/ (RC1+RC2+RC3)

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
29	Past year but not past month, or past 3 years but not past year				33	(SLT3-4,9,17)	1. R2/(R2+R3)
30		Past year but not past month, or past 3 years but not past year			43	(SLT3-4,10,18)	1. R2/(R2+R3)
31	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			10	(SLT3-4,17-18)	1. R2/(R2+R3)
32	Past year but not past month, or past 3 years but not past year	(Past month)		Missing	0	(SLT3-4,6,8,17)	1. R2/(R2+R3) 2. DS 3. PMS
33	(Past month)	Past year but not past month, or past 3 years but not past year	Missing		0	(SLT3-5,7,18)	1. R2/(R2+R3) 2. DC 3. PMC

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
34	Past year but not past month, or past 3 years but not past year	Missing (lifetime use known)		Missing	0	(SLT3-4,6,17)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
34	Past year but not past month, or past 3 years but not past year	Missing (lifetime use imputed)		Missing	0		
35	Missing (lifetime use known)	Past year but not past month, or past 3 years but not past year	Missing		0	(SLT3-5,18)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
35	Missing (lifetime use imputed)	Past year but not past month, or past 3 years but not past year	Missing		0		
36	Past year but not past month, or past 3 years but not past year	Not past year			0	(SLT3-4,14,17)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
37	Not past year	Past year but not past month, or past 3 years but not past year			1	(SLT3-4,13,18)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

Table H.9 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
38	Past 3 years	Past 3 years	Missing	Missing	0	(SLT1-6,11-12)	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)
39	Not past month	Past year but not past month, or past 3 years but not past year			1	(SLT3-4,15,18)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
40	Missing (lifetime use known)	Past 3 years	Missing	Missing	0	(SLT2-6)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS/ (RS1+RS2+RS3) 7. RS1*(1-DS)*PMS/ (RS1+RS2+RS3)
40	Missing (lifetime use imputed)	Past 3 years	Missing	Missing	0		
	Lifetime user, nothing missing				11,682		
	Imputed to lifetime nonuse				32		
	Lifetime nonuser, nothing missing				56,268		

[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. RC1 = P(past month chewing tobacco use | lifetime chewing tobacco use)
5. RC2 = P(past year but not past month chewing tobacco use | lifetime chewing tobacco use)
6. RS1 = P(past month snuff use | lifetime snuff use)
7. RS2 = P(past year but not past month snuff use | lifetime snuff use)
8. DC = P(daily chewing tobacco use | past month chewing tobacco use)
9. DS = P(daily snuff use | past month snuff use)
10. PMC = P(chewing tobacco use on a given day in the past month | past month use of chewing tobacco)
11. PMS = P(snuff use on a given day in the past month | past month use of snuff)

Table H.10 Pipe User Restrictions

Missingness Pattern		Number of Cases	Constraints
#	Recency		
1	Missing (lifetime use known)	4	(None)
1	Missing (lifetime use imputed)	0	(None)
	Lifetime user, nothing missing	6,181	
	Imputed to lifetime nonuse	10	
	Lifetime nonuser, nothing missing	62,113	

NOTE: For pipes, only a two-level recency-of-use variable was imputed. The imputation was univariate, both in terms of the predictive mean vector and the final assignment. Item nonrespondents were handled identically, whether or not lifetime use was imputed.

Table H.11 Constraints for Various Drugs

Drug	Constraint #	Constraint
Alc, Mrj, Inh, Her, Trn, Sed	C1	Donor's proportion of past year use * recipient's max number of days could have used in past year must be less than (or equal to) the recipient's maximum possible past year frequency of use. The recipient's maximum possible 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 the recipient could have used, as determined by the month of first use (2) if the maximum period the recipient could have used is greater than 30, but the recipient is a past month user with a nonmissing 30-day frequency, the past year frequency must be less than or equal to the maximum period (minus the number of days the recipient didn't use in the past month) (3) if the recipient is not a past month user, the past year frequency must be less than or equal to the maximum period (minus 30)
Alc, Mrj, Inh, Her	C2	Donor's proportion of past year use * recipient's min number of days could have used in past year must be greater than (or equal to) the recipient's minimum possible past year frequency of use. The recipient's minimum possible frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month user, it must be at least as much as the 30-day frequency (2) if the recipient is not a past month user but a past year user, it must be at least 1
Alc, Mrj, Inh, Her, Trn, Sed	C3	(Recipient's proportion of past year use * max number of days could have used in past year) must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Alc, Mrj, Inh	C4	(Donor's proportion of past year use * recipient's number of days could have used in past year) must be greater than or equal to 30-day use
Alc, Mrj, Inh, Her	C5	Donor's 30-day use must be less than number of days between recipient's interview date and birthday (inclusive)
Alc, Mrj, Inh, Her	C6	Donor's 30-day use must be less than the recipient's maximum number of days could have used in past 30 days
Alc, Mrj, Inh, Her	C7	Donor's 30-day use must be greater than the recipient's minimum number of days could have used in past 30 days
Alc	C8	Donor's 30-day use must be greater than recipient's DR5DAY (# days had 5+ drinks in past 30 days)
Alc, Mrj, Inh, Her	C9	Donor's 30-day use must be greater than (donor's proportion of past year use * recipient's max number of days could have used in past year [minus 335])
Alc, Mrj, Inh, Her, Trn, Sed	C10	Donor must be a past month user (recency = 1)
Alc, Mrj, Inh, Her	C11	If recipient's age at first use equals his or her current age, the donor's 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first drug use (inclusive) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)

Table H.11 Constraints for Various Drugs (continued)

Drug	Constraint #	Constraint
Alc, Mrj, Inh, Her	C12	If recipient's age at first use equals his or her current age, (1) recipient's donor's proportion of past year use * recipient's max number of days could have used in 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 use * recipient's max number of days could have used in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Alc, Mrj, Inh, Her	C13	Recipient's estimated 30-day frequency is not given/legitimately skipped (estimated frequency not equal to 1-6)
Alc, Mrj, Inh, Her	C14	If recipient's age at first use equals his or her current age, (1) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (minus 29) and (2) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than the recipient's days between the interview date and birthday (minus 29)
Alc, Mrj, Inh, Her, Trn, Sed	C15	Donor must be a past year (but not past month) user (recency = 2)
Alc	C16	Donor's DR5DAY value is less than or equal to recipient's 30-day frequency
Alc	C17	If recipient's age at first 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)
Alc, Mrj, Inh, Her, Trn, Sed	C18	Donor must be a past month or past year (but not past month) user (recency = 1 or 2)
Alc, Mrj, Inh, Her	C19	Donor's proportion of past year use * recipient's max number of days could have used in past year must be greater or equal to donor's 30-day frequency
Alc, Mrj, Inh, Her	C20	If recipient's age at first use equals his or her current age, (1) donor's proportion of past year used * recipient's max number of days could have used in 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 used * recipient's max number of days could have used in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Alc, Mrj, Inh, Her	C21	If recipient's month/year of first use data indicate that he/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 use * recipient's max number of days could have used in past year must be greater than recipient's 30-day frequency
Alc	C22	If recipient's month/year of first use data indicate that he/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 use * recipient's max number of days could have used in past year must be greater than recipient's DR5DAY value

Table H.11 Constraints for Various Drugs (continued)

Drug	Constraint #	Constraint
Alc, Mrj, Inh, Her	C23	If recipient's month/year of first use data indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than donor's 30-day frequency
Alc, Mrj, Inh, Her	C24	If recipient's month/year of first use data indicate that he/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 use * recipient's max number of days could have used in past year must be greater than donor's 30-day frequency
Trn, Sed	C25	If recipient's month/year of first use data indicate that he/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 use * recipient's max number of days could have used in past year must be greater than 1
Trn, Sed	C26	If recipient's month/year of first use data indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than 1

Table H.12 Restrictions and Portion of the Predictive Mean Vector for Alcohol Users

Missingness Pattern							
#	Recency	12-Month Freq.	30-Day Freq.	30-Day Binge Drink	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	(Past month)	Missing	Missing		27	(C1-13,22)	1. PM 2. PY
2	(Past month)		Missing		221	(C5-8,10-11,13)	1. PM
3	(Past month)	Missing			173	(C1-4,10,12,21)	1. PY
4	(Past year but not past month)	Missing			150	(C1-3,14-15)	1. PY
5	(Past month)			Missing	580	(C10,16,17)	1. PMB
6	(Past month)		Missing	Missing	22	(C5-7,10-11,13)	1. PM 2. PMB
7	(Past month)	Missing		Missing	73	(C1-4,10,12,16-17,21)	1. PY 2. PMB
8	(Past month)	Missing	Missing	Missing	15	(C1-7,9-13,24)	1. PM 2. PY 3. PMB
9	Past Year		Missing	Missing	354	(C5-7,11,13,15)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)
10	Past year	Missing	Missing	Missing	82	(C1-3,5-9,11-14,18,23)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)
11	Lifetime (known)	Missing	Missing	Missing	459	(C1-7,9,11-14)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB
11	Lifetime (imputed)	Missing	Missing	Missing	6		
	(30-day binge drink response missing in raw data. Logically set to zero based on responses in other parts of questionnaire. No other responses missing.)				65		
	Lifetime user, nothing missing				47,571		
	Imputed to lifetime nonuse				10		
	Lifetime nonuser, nothing missing				18,500		

[1] The predictive mean vector components are defined by the following:

1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)
5. PMB = P(binge drinking on a given day in the past month | past month use)

Table H.13 Restrictions and Portion of the Predictive Mean Vector for Users of Marijuana, Inhalants, and Heroin

Missingness Pattern						
#	Recency	12-Month Freq.	30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
1	(Past month)	Missing	Missing	Marijuana: 9	(C1-7,9-13,24)	1. PM 2. PY
				Inhalants: 7		
				Heroin: 0		
2	(Past month)		Missing	Marijuana: 11	(C5-7,10-11,13)	1. PM
				Inhalants: 2		
				Heroin: 0		
3	(Past month)	Missing		Marijuana: 78	(C1-4,10,12,21)	1. PY
				Inhalants: 19		
				Heroin: 3		
4	(Past year but not past month)	Missing		Marijuana: 48	(C1-3,13-14)	1. PY
				Inhalants: 43		
				Heroin: 2		
5	Past year		Missing	Marijuana: 89	(C5-7,11,13,18)	1. R1/(R1+R2) 2. R1* PM/(R1+R2)
				Inhalants: 15		
				Heroin: 2		
6	Past year	Missing	Missing	Marijuana: 109	(C1-3,5-7,9,11-14,18-19,23)	1. R1/(R1+R2) 2. R1* PM/(R1+R2) 3. PY
				Inhalants: 8		
				Heroin: 13		
7	Missing (lifetime use known)	Missing	Missing	Marijuana: 228	(C1-3,5-7,9,11-14,19,20)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
				Inhalants: 285		
				Heroin: 6		
7	Missing (lifetime use imputed)	Missing	Missing	Marijuana: 14		
				Inhalants: 9		
				Heroin: 0		

Table H.13 Restrictions and Portion of the Predictive Mean Vector for Users of Marijuana, Inhalants, and Heroin (continued)

Missingness Pattern		Number of Cases	Constraints	Predictive Mean Vector[1]
	Lifetime user, nothing missing	Marijuana: 26,429		
		Inhalants: 7,487		
		Heroin: 810		
	Imputed to lifetime nonuse	Marijuana: 23		
		Inhalants: 118		
		Heroin: 37		
	Lifetime nonuser, nothing missing	Marijuana: 41,270		
		Inhalants: 60,315		
		Heroin: 67,435		

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Table H.14 Restrictions and Portion of the Predictive Mean Vector for Users of Tranquilizers and Sedatives

Missingness Pattern			Number of Cases	Constraints	Predictive Mean Vector[1]
#	**Recency**	**12-Month Frequency**			
1	(Past month)	Missing	Tranquilizers: 14	(C1,3,10,25)	1. PY
			Sedatives: 2		
2	(Past year but not past month)	Missing	Tranquilizers: 14	(C1,3,15)	1. PY
			Sedatives: 7		
3	Past year		Tranquilizers: 0	(C18)	1. R1/(R1+R2)
			Sedatives: 2		
4	Past year	Missing	Tranquilizers: 9	(C1,3,18,26)	1. R1/(R1+R2) 2. PY
			Sedatives: 2		
5	Missing (lifetime use known)	Missing	Tranquilizers: 130	(C1,3,18)	1. R1 2. R2 3. (R1+R2)*PY
			Sedatives: 30		
5	Missing (lifetime use imputed)	Missing	Tranquilizers: 3		
			Sedatives: 4		
	Lifetime user, nothing missing		Tranquilizers: 5,611		
			Sedatives: 1,569		
	Imputed to lifetime nonuse		Tranquilizers: 131		
			Sedatives: 157		
	Lifetime nonuser, nothing missing		Tranquilizers: 62,396		
			Sedatives: 66,535		

NOTE: The missingness patterns and predictive mean vectors for the tranquilizer and sedatives modules were identical.

[1]The predictive mean vector components are defined by the following:

1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

Table H.15 Constraints for Cocaine and Crack

Constraint #	Constraint
Coc1	Donor must be a past month cocaine user (cocaine recency = 1)
Coc2	Donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year must be less than or equal to the recipient's maximum possible past year cocaine frequency of use. The recipient's maximum possible cocaine frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used cocaine, as determined by the month of first use (2) if the maximum period the recipient could have used cocaine is greater than 30, but the recipient is a past month cocaine user with a nonmissing 30-day frequency, the past year cocaine frequency must be less than or equal to the maximum period (the number of days the recipient did not use in the past month) (3) if the 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)
Coc3	Donor's proportion of past year cocaine use * recipient's min number of days could have used cocaine in past year must be greater than or equal to the recipient's minimum possible past year cocaine frequency of use. The recipient's minimum possible cocaine frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month cocaine user, it must be at least as much as the 30-day frequency (2) if the recipient is not a past month cocaine user but a past year cocaine user, it must be at least 1
Coc4	Recipient's proportion of past year cocaine use * max number of days could have used cocaine in past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Coc5	Donor's proportion of past year cocaine use * recipient's number of days could have used cocaine in past year must be greater than or equal to 30-day use
Coc6	Donor's 30-day cocaine use must be less than number of days between recipient's interview date and birthday (inclusive)
Coc7	Donor's 30-day cocaine use must be less than the recipient's maximum number of days could have used in past 30 days
Coc8	Donor's 30-day cocaine use must be greater than the recipient's minimum number of days could have used in past 30 days
Coc9	If recipient's age at first cocaine use equals his or her current age, the donor's cocaine 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first cocaine use (inclusive) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)

Table H.15 Constraints for Cocaine and Crack (continued)

Constraint #	Constraint
Coc10	If recipient's age at first cocaine use equals his or her current age, (1) recipient's donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in 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 max number of days could have used cocaine in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Coc11	Recipient's estimated cocaine 30-day frequency is not given/legitimately skipped (estimated cocaine frequency not equal to 1-6)
Coc12	Donor's crack recency equals recipient's crack recency
Coc13	Donor must be a past year (but not past month) cocaine user (cocaine recency = 2)
Coc14	If recipient's age at first cocaine use equals his or her current age, donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year cannot be greater than recipient's days between his or her interview date and date of first cocaine use (minus 29)
Coc15	Donor must be a past month or past year (but not past month) cocaine user (cocaine recency = 1 or 2)
Coc16	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)
Coc17	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 cannot equal 3)
Coc18	Donor's proportion of past year crack use * recipient's max number of days could have used crack in past year must be less than (or equal) the recipient's maximum possible past year crack frequency of use The 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 than the maximum period the recipient could have used crack, as determined by the month of first use (2) if the maximum period the recipient could have used crack is greater than 30, but the recipient is a past month crack user with a nonmissing 30-day frequency, the past year crack frequency must be less than or equal to the maximum period (the number of days the recipient did not use in the past month) (3) if the 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)
Coc19	Donor's proportion of past year crack use * recipient's min number of days could have used crack in past year must be greater than (or equal to) the recipient's minimum possible past year crack frequency of use The recipient's minimum possible crack frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month crack user, it must be at least as much as the 30-day frequency (2) if the recipient is not a past month crack user but a past year crack user, it must be at least 1

Table H.15 Constraints for Cocaine and Crack (continued)

Constraint #	Constraint
Coc20	(Recipient's proportion of past year crack use * max number of days could have used crack in past year) must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Coc21	(Donor's proportion of past year crack use * recipient's number of days could have used crack in past year) must be greater than or equal to 30-day use
Coc22	Donor's 30-day crack use must be less than number of days between recipient's interview date and birthday (inclusive)
Coc23	Donor's 30-day crack use must be less than the recipient's maximum number of days could have used in past 30 days
Coc24	Donor's 30-day crack use must be greater than the recipient's minimum number of days could have used in past 30 days
Coc25	If recipient's age at first crack use equals his or her current age, the donor's crack 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first crack use (inclusive) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Coc26	If recipient's age at first crack use equals his or her current age, (1) recipient's donor's proportion of past year crack use * recipient's max number of days could have used crack in 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 crack use * recipient's max number of days could have used crack in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Coc27	Recipient's estimated 30-day crack frequency is not given/legitimately skipped (estimated crack frequency not equal to 1-6)
Coc28	Donor must be a past month crack user (crack recency = 1)
Coc29	Donor must be a past month or past year (not past month) crack user (crack recency = 1, 2)
Coc30	Donor must be a past month, past year (not past month), or lifetime (but not past year) crack user (crack recency = 1, 2)
Coc31	Donor's cocaine recency must equal recipient's cocaine recency, or donor's cocaine recency must equal recipient's cocaine recency (minus 10)
Coc32	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 cannot equal 3)
Coc33	Donor must be a past year (but not past month) crack user (crack recency = 2)
Coc34	If recipient's age at first crack use equals his or her current age, donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than recipient's days between his or her interview date and date of first crack use (minus 29)
Coc35	If recipient's month/year of first use data for cocaine indicate that he/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 cocaine use * recipient's max number of days could have used cocaine in past year must be greater than recipient's cocaine 30-day frequency

Table H.15 Constraints for Cocaine and Crack (continued)

Constraint #	Constraint
Coc36	If recipient's month/year of first use data for cocaine indicate that he/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 cocaine use * recipient's max number of days could have used cocaine in past year must be greater than donor's cocaine 30-day frequency
Coc37	If recipient's month/year of first use data for cocaine indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month cocaine user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than donor's cocaine 30-day frequency
Coc38	If recipient's month/year of first use data for crack indicate that he/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 crack use * recipient's max number of days could have used crack in past year must be greater than recipient's crack 30-day frequency
Coc39	If recipient's month/year of first use data for crack indicate that he/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 crack use * recipient's max number of days could have used crack in past year must be greater than donor's crack 30-day frequency
Coc40	If recipient's month/year of first use data for crack indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month crack user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than donor's crack 30-day frequency

Table H.16 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
1	(Past month)		Missing		Missing		15	(Coc1-12,36)	1. PM 2. PY
2	(Past month)				Missing		17	(Coc1,6-9,11-12)	1. PM
3	(Past month)		Missing				5	(Coc2-4,10,12,35)	1. PY
4	(Past year not past month)		Missing				39	(Coc2-4,12-14)	1. PY
5	Past year				Missing		24	(Coc6-9,11-12,15)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year		Missing		Missing		13	(Coc2-12,15,37)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing (lifetime use known)		Missing		Missing		95	(Coc2-12,16-17)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
7	Missing (lifetime use imputed)		Missing		Missing		1		
8	(Past month)	(Past month)		Missing		Missing	0	(Coc1,18-27,39)	1. PM 2. PY
9	(Past month)	(Past month)				Missing	0	(Coc1,22-25,27-28)	1. PM
10	(Past month)	(Past month)		Missing			0	(Coc15,18-Coc20,26,28,38)	1. PY
11	(Past year not missing)	(Past year not past month)		Missing			0	(Coc15,18-Coc20,26,29)	1. PY

Table H.16 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
12	(Past month)	Past year				Missing	3	(Coc1,22-25,27,29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
13	(Past month)	Past year		Missing		Missing	1	(Coc1,18-27,29,40)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
14	(Past month)	Missing (Lifetime use known)		Missing		Missing	6	(Coc16,18-26,30-32)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
14	(Past month)	Missing (Lifetime use imputed)		Missing		Missing	1		
15	(Past month)	(Past month)	Missing	Missing			0	(Coc1-4,10,18-20,26,28,35,38)	1. PY
16	(Past month)	(Past year but not past month)	Missing	Missing			0	(Coc1-4,10,18-20,26,33,35)	1. PY
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			2	(Coc2-4,14,18-20,33-34)	1. PY
18	(Past month)	(Past month)			Missing	Missing	0	(Coc1,6-9,11,22-25,27-28)	1. PM
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	1	(Coc1-11,18-28,36,39)	1. PM 2. PY
20	(Past month)	(Past month)	Missing		Missing	Missing	0	(Coc1-11,16,22-25,27-28,36)	1. PM 2. PY
21	(Past month)	(Past month)		Missing	Missing	Missing	0	(Coc1,6-9,11,18-28,39)	1. PM 2. PY

Table H.16 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.			
22	(Past month)	(Past month)	Missing	Missing	Missing		0	(Coc1-11,18-21,26,28,36,38)	1. PM 2. PY
23	(Past month)	(Past year not past month)	Missing	Missing	Missing		0	(Coc1-11,18-20,33,34,36)	1. PM 2. PY
24	(Past month)	(Past month)	Missing	Missing		Missing	0	(Coc1-4,10,18-26,28,36)	1. PM 2. PY
25	(Past month)	(Past month)		Missing	Missing		0	(Coc1,6-9,18-20,26,28,38)	1. PM 2. PY
26	(Past month)	(Past year not past month)		Missing	Missing		0	(Coc1,6-9,11,18-20,26,33)	1. PM 2. PY
27	(Past month)	(Past month)	Missing			Missing	0	(Coc1-4,10,22-25,27-28,35)	1. PM 2. PY
28	Past year	Past year			Missing	Missing	2	(Coc6-9,11,15,22-25,27,29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
29	Past year	Past year	Missing		Missing	Missing	1	(Coc2-11,15,21-25,27,29,37)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
30	Past year	Past year		Missing	Missing	Missing	9	(Coc6-9,11,15,18-27,29,40)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
31	Past year	Past year	Missing	Missing	Missing	Missing	4	(Coc2-11,15,18-27,29,37,40)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

Table H.16 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
32	Past year	Missing (lifetime use known)		Missing	Missing	Missing	9	(Coc1,6-9,11,15,18-27,30)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
32	Past year	Missing (lifetime use imputed)		Missing	Missing	Missing	0		
33	Past year	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	(Coc2-11,15,18-27,30,32,37)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
33	Past year	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0		
34	(Past month)	Missing (lifetime use known)		Missing	Missing	Missing	0	(Coc1,6-9,11,18-27,30,32)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
34	(Past month)	Missing (lifetime use imputed)		Missing	Missing	Missing	0		
35	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	(Coc1-11,18-27,30,36)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
35	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0		

Table H.16 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	12	(Coc2-11,16-27,30)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
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	3		
37	(Past Month)	Past year		Missing	Missing	Missing	0	(Coc1,6-9,11,18-27,29,40)	1. R1/(R1 + R2) 2. PM 3. PY[2]
38	(Past Month)	Past year			Missing	Missing	1	(Coc1,6-9, 22-25,27,29)	1. R1/(R1+R2) 2. PM[3]
	Lifetime user, nothing missing						7,787		
	Imputed to lifetime nonuse						16		
	Lifetime nonuser, nothing missing						60,241		

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); and 4. PY = P(cocaine use on a given day in the past year | past year use of cocaine).

[2]This predictive mean vector will be changed for 2006 processing, to the following: 1. R1/(R1+R2); 2. R1*PM/(R1+R2); and 3. PY. See Section 6.5.5.2 for details.

[3]This predictive mean vector will be changed for 2006 processing, to the following: 1. R1/(R1+R2) and 2. R1*PM/(R1+R2). See Section 6.5.5.2 for details.

Table H.17 Constraints for Hallucinogens (Including LSD, PCP, and Ecstasy)

Constraint #	Constraint
Hal1	Donor's proportion of past year hallucinogen use * recipient's max number of days could have used hallucinogens in past year must be less than or equal to the recipient's maximum possible past year hallucinogen frequency of use. The recipient's maximum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used hallucinogens, as determined by the month of first use (2) if the maximum period the recipient could have used hallucinogens is greater than 30, but the recipient is a past month 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 the recipient did not use hallucinogens in the past month) (3) if the 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)
Hal2	Donor's proportion of past year hallucinogen use * recipient's min number of days could have used hallucinogens in past year must be greater than (or equal to) the recipient's minimum possible past year hallucinogen frequency of use. The recipient's minimum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month hallucinogen user, it must be at least as much as the hallucinogen 30-day freq (2) if the recipient is not a past month hallucinogen user but a past year hallucinogen user, it must be at least 1
Hal3	(Recipient's proportion of past year hallucinogen use * max number of days could have used hallucinogens in past year) must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Hal4	Donor's 30-day hallucinogen use must be less than number of days between recipient's interview date and birthday (inclusive)
Hal5	Donor's 30-day hallucinogen use must be less than the recipient's maximum number of days could have used hallucinogens in past 30 days
Hal6	Donor's 30-day hallucinogen use must be greater than the recipient's minimum number of days could have used hallucinogens in past 30 days
Hal7	Donor must be an LSD user (LSD recency not equal to 91)
Hal8	Donor must be a PCP user (PCP recency not equal to 91)
Hal9	Donor must be an Ecstasy user (Ecstasy recency not equal to 91)
Hal10	Donor's LSD recency must equal recipient's LSD recency
Hal11	Donor's PCP recency must equal recipient's PCP recency
Hal12	Donor's Ecstasy recency must equal recipient's Ecstasy recency
Hal13	Donor must be an LSD and PCP user (LSD and PCP recencies not equal to 91)
Hal14	Donor must be an LSD and Ecstasy user (LSD and Ecstasy recencies not equal to 91)
Hal15	Donor must be a PCP and Ecstasy user (PCP and Ecstasy recencies not equal to 91)

Table H.17 Constraints for Hallucinogens (Including LSD, PCP, and Ecstasy) (continued)

Constraint #	Constraint
Hal16	Donor must be an LSD and PCP and Ecstasy user (LSD and PCP and Ecstasy recencies not equal to 91)
Hal17	Donor's must be a past month hallucinogens user (hallucinogen recency = 1)
Hal18	Donor must be a hallucinogen past year (but not past month) or past month user (hallucinogen recency = 1 or 2)
Hal19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
Hal20	Donor must be an LSD past year (but not past month) or past month user (LSD recency = 1 or 2)
Hal21	Donor must be a PCP past year (but not past month) or past month user (PCP recency = 1 or 2)
Hal22	Donor must be an Ecstasy past year (but not past month) or past month user (Ecstasy recency = 1 or 2)
Hal23	Donor must not be an LSD past year (but not past month) or past month user (LSD recency not equal to 1 or 2)
Hal24	Donor must not be a PCP past year (but not past month) or past month user (PCP recency not equal to 1 or 2)
Hal25	Donor must not be an Ecstasy past year (but not past month) or past month user (Ecstasy recency not equal to 1 or 2)
Hal26	Donor's hallucinogen recency must equal recipient's hallucinogen recency, or donor's hallucinogen recency must equal recipient's hallucinogen recency (minus 10)
Hal27	If recipient's month/year of first use data for hallucinogens indicate that he/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 hallucinogens use * recipient's max number of days could have used hallucinogens in past year must be greater than recipient's hallucinogens 30-day frequency
Hal28	If recipient's month/year of first use data for hallucinogens indicate that he/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 hallucinogens use * recipient's max number of days could have used hallucinogens in past year must be greater than donor's hallucinogens 30-day frequency
Hal29	If recipient's month/year of first use data for hallucinogens indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month hallucinogens user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than donor's hallucinogens 30-day frequency
Hal30	If recipient is a past month hallucinogens user and recipient's month/year of first use data for hallucinogens indicate that he/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 hallucinogens use * recipient's max number of days could have used hallucinogens in past year must be greater than donor's hallucinogens 30-day frequency
Hal31	If recipient has never used hallucinogens other than LSD, PCP, and Ecstasy, than donor must not have recency values that would cause the recipient to have imputation revised recency for overall hallucinogens less than the minimum of the imputation revised recencies for LSD, PCP, and Ecstasy

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
1		Missing (lifetime use known)					2	(Hal7,11-12,26)	1. R1 2. R2
1		Missing (lifetime use imputed)					0		
2			Missing (lifetime use known)				2	(Hal8,10,12,26)	1. R1 2. R2
2			Missing (lifetime use imputed)				0		
3		Missing (lifetime use known)	Missing (lifetime use known)				1	(Hal7-8,12,26,31)	1. R1 2. R2
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		
4	(Past month)				Missing	Missing	1	(Hal1-6,17,28)	1. PM 2. PY
5	(Past month)					Missing	9	(Hal4-6,17)	1. PM
6	(Past year)				Missing		14	(Hal1-3,18,30)	1. PY
7	(Past month)	Missing (lifetime use known)				Missing	0	(Hal4-7,11-12,17)	1. R1 2. R2 3. PM

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
7	(Past month)	Missing (lifetime use imputed)				Missing	0		
8	(Past month)		Missing (lifetime use known)			Missing	0	(Hal4-6,8,10,12,17)	1. R1 2. R2 3. PM
8	(Past month)		Missing (lifetime use imputed)			Missing	0		
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)			Missing	0	(Hal4-8,12,17,31)	1. R1 2. R2 3. PM
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	(Hal1-3,7,11-12,18,30)	1. R1 2. R2 3. PY
10	(Past year)	Missing (lifetime use imputed)			Missing		0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
11	(Past year)		Missing (lifetime use known)		Missing		0	(Hal1-3,8,10,12,18,30)	1. R1 2. R2 3. PY
11	(Past year)		Missing (lifetime use imputed)		Missing		0		
12	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)		Missing		0	(Hal1-3,7-8,12,18,30,31)	1. R1 2. R2 3. PY
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	(Hal1-6-7,11-12,17,28)	1. R1 2. R2 3. PM 4. PY
13	(Past month)	Missing (lifetime use imputed)			Missing	Missing	0		
14	(Past month)		Missing (lifetime use known)		Missing	Missing	0	(Hal1-6,8,10,12,17,28)	1. R1 2. R2 3. PM 4. PY
14	(Past month)		Missing (lifetime use imputed)		Missing	Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	Missing	0	(Hal1-8,12,17,28,31)	1. R1 2. R2 3. PM 4. PY
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	17	(Hal4-6,10-12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
17	Past year	(Not past month)	(Not past month)	(Not past month)	Missing	Missing	7	(Hal1-6,10-12,18,29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
18	Past year	Past year	(Not past month)	(Not past month)		Missing	11	(Hal4-6,11-12,18,20,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
19	Past year	(Not past month)	Past year	(Not past month)		Missing	2	(Hal4-6,10,12,18,21,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
20	Past year	Past year	Past year	(Not past month)		Missing	0	(Hal4-6,12,18,20-21,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
21	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)		Missing	10	(Hal4-7,11-12,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
21	Past year	Missing (lifetime use imputed)	(Not past month)	(Not past month)		Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
22	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)		Missing	1	(Hal4-6,8,10,12,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
22	Past year	(Not past month)	Missing (lifetime use imputed)	(Not past month)		Missing	0		
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)		Missing	3	(Hal4-8,12,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
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	3	(Hal1-6,11-12,18,20,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
25	Past year	(Not past month)	Past year	(Not past month)	Missing	Missing	1	(Hal1-6,10,12,18,21,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
26	Past year	Past year	Past year	(Not past month)	Missing	Missing	0	(Hal1-6,12,18,20,21,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
27	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)	Missing	Missing	0	(Hal1-6,7,11-12,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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	1	(Hal1-6,8,11-12,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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	(Hal1-8,12,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
30	Missing (lifetime use known)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	71	(Hal1-6,10-12,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
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	39	(Hal1-7,11-12,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
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	2		
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	20	(Hal1-6,8,10,12,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
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	4	(Hal1-8,12,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
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)			4	(Hal9-11,26)	1. R1 2. R2
34				Missing (lifetime use imputed)			0		
35		Missing (lifetime use known)		Missing (lifetime use known)			1	(Hal7,9,11,26,31)	1. R1 2. R2
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	**Hallucinogen Recency**	**LSD Recency**	**PCP Recency**	**Ecstasy Recency**	**Hallucinogen 12-Mo. Freq.**	**Hallucinogen 30-Day Freq.**	**Number of Cases**	**Constraints**	**Predictive Mean Vector[1]**
36			Missing (lifetime use known)	Missing (lifetime use known)			1	(Hal8-10,26,31)	1. R1 2. R2
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	(Hal7-9,26,31)	1. R1 2. R2
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
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	(Hal4-6,9-11,17)	1. R1 2. R2 3. PM
38	(Past month)			Missing (lifetime use imputed)		Missing	0		
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)		Missing	0	(Hal4-7,9,11,17,31)	1. R1 2. R2 3. PM
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use imputed)		Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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	(Hal4-6,8-10,17,31)	1. R1 2. R2 3. PM
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	(Hal4-9,17,31)	1. R1 2. R2 3. PM
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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	(Hal1-3,9-11,18,30)	1. R1 2. R2 3. PY
42	(Past year)			Missing (lifetime use imputed)	Missing		0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use known)	Missing		0	(Hal1-3,7,9,11,18,30,31)	1. R1 2. R2 3. PY
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	(Hal1-3,8-10,18,30,31)	1. R1 2. R2 3. PY
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	**Hallucinogen Recency**	**LSD Recency**	**PCP Recency**	**Ecstasy Recency**	**Hallucinogen 12-Mo. Freq.**	**Hallucinogen 30-Day Freq.**	**Number of Cases**	**Constraints**	**Predictive Mean Vector[1]**
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	(Hal1-3,7-9,18,30,31)	1. R1 2. R2 3. PY
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
46	(Past month)			Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,9-11,17,28)	1. R1 2. R2 3. PM 4. PY
46	(Past month)			Missing (lifetime use imputed)	Missing	Missing	0		
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)	Missing	Missing	0	(Hal1-7,9,11,17,28,31)	1. R1 2. R2 3. PM 4. PY
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	(Hal1-6,8-10,17,28,31)	1. R1 2. R2 3. PM 4. PY
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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	(Hal1-9,17,28,31)	1. R1 2. R2 3. PM 4. PY
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 H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
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	1	(Hal4-6,10-11,18,22,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
51	Past year	Past year	(Not past month)	Past year		Missing	1	(Hal4-6,11,18,20,22,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
52	Past year	(Not past month)	Past year	Past year		Missing	0	(Hal4-6,10,18,21-22,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
53	Past year	Past year	Past year	Past year		Missing	0	(Hal4-6,18,20-22,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)		Missing	5	(Hal4-6,9-11,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)		Missing	0		
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)		Missing	0	(Hal4-7,9,11,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use imputed)		Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
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	2	(Hal4-6,8-10,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
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		
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	(Hal4-9,18,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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		
58	Past year	(Not past month)	(Not past month)	Past year	Missing	Missing	1	(Hal1-6,10-11,18,22,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
59	Past year	Past year	(Not past month)	Past year	Missing	Missing	1	(Hal1-6,11,18,20,22,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
60	Past year	(Not past month)	Past year	Past year	Missing	Missing	0	(Hal1-6,10,18,21-22,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
61	Past year	Past year	Past year	Past year	Missing	Missing	0	(Hal1-6,18,20-22,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,9-11,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0		
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-7,9,11,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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	(Hal1-6,8-10,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
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		
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-9,18,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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		
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	39	(Hal1-6,9-11,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
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	7	(Hal1-7,9,11,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	1		
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		
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	(Hal1-6,8-10,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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		
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	(Hal1-9,19,31)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
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		
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		

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	**Hallucinogen Recency**	**LSD Recency**	**PCP Recency**	**Ecstasy Recency**	**Hallucinogen 12-Mo. Freq.**	**Hallucinogen 30-Day Freq.**	**Number of Cases**	**Constraints**	**Predictive Mean Vector[1]**
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		
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			1	(Hal10-11,22,26)	1. R1/(R1+R2)
71		Past year	Past year				0	(Hal12,20-21,26,31)	1. R1/(R1+R2)
72		Past year		Past year			0	(Hal11,20,22,26,31)	1. R1/(R1+R2)

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
73	(Past month)	Past year			Missing	Missing	0	(Hal1-6,11-12,17,20,28)	1. R1/(R1+R2) 2. PM 3. PY
74	(Past month)		Past year		Missing	Missing	0	(Hal1-6,10,12,17,21,28)	1. R1/(R1+R2) 2. PM 3. PY
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use known)	Missing	Missing	0	(Hal1-7,9,18,21,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
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	(Hal1-6,9-10,18,21,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
76	Past year	(Not past month)	Past year	Missing (lifetime use imputed)	Missing	Missing	0		
77			Past year				0	(Hal10,12,21,26)	1. R1/(R1+R2)
78		Past year					1	(Hal11-12,20,26)	1. R1/(R1+R2)

Table H.18 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and Ecstasy) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
79	(Past month)	Past year				Missing	0	(Hal4-7,11-12,17)	1. R1/(R1+R2) 2. PM
80	(Past month)		Past year			Missing	1	(Hal4-6,8,10,12,17)	1. R1/(R1+R2) 2. PM
81	(Past month)			Past year		Missing	0	(Hal4-6,9-11,17)	1. R1/(R1+R2) 2. PM
82	(Past month)	Past year			Missing		0	(Hal1-3,7,11-12,17,27)	1. R1/(R1+R2) 2. PY
83	(Past month)		Past year		Missing		0	(Hal1-3,8,10,12,17,27)	1. R1/(R1+R2) 2. PY
84	(Past month)			Past year	Missing		0	(Hal1-3,9-11,17,27)	1. R1/(R1+R2) 2. PY
85	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Past year		Missing	0	(Hal4-6,13,18,22,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
86	Past year	Missing (lifetime use known)	Past year		Missing	Missing	0	(Hal1-7,12,18,21,29,31)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
	Lifetime user, nothing missing						9,387		
	Imputed to lifetime nonuse						205		
	Lifetime nonuser, nothing missing						58,423		

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 use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Table H.19 Constraints for Stimulants, Methamphetamines, Pain Relievers, and OxyContin

Constraint #	Constraint
Stm1	Donor's proportion of past year parent drug use * recipient's max number of days could have used parent drug in past year must be less than (or equal to) the recipient's maximum possible past year parent drug frequency of use. The 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 the recipient could have used parent drug, as determined by the month of first use (2) if the 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)
Stm2	Donor's proportion of past year parent drug use * recipient's min number of days could have used parent drug in past year must be greater than (or equal to) the 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.)
Stm3	(Recipient's proportion of past year parent drug use * max number of days could have used parent drug in past year) must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Stm4	Donor must be a past month parent drug user (parent drug recency = 1)
Stm5	If recipient's age at first parent drug use equals his or her current age, (1) recipient's donor's proportion of past year parent drug use * recipient's max number of days could have used parent drug in 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 parent drug use * recipient's max number of days could have used parent drug in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)
Stm6	Donor must be a past year (but not past month) parent drug user (parent drug recency = 2)
Stm7	If recipient's age at first parent drug use equals his or her current age, (1) recipient's donor's proportion of past year parent drug use* recipient's max number of days could have used parent drug in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (minus 29) and (2) donor's proportion of past year parent drug use * recipient's max number of days could have used parent drug in past year cannot be greater than the recipient's days between his or her interview date and birthday (minus 29)
Stm8	Donor must be a past month or past year (but not past month) parent drug user (parent drug recency = 1 or 2)
Stm9	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).

Table H.19 Constraints for Stimulants, Methamphetamines, Pain Relievers, and OxyContin (continued)

Constraint #	Constraint
Stm10	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)
Stm11	If the number of days between the recipient's interview and birthday (inclusive) is between 0 and 30, child drug recency must not equal 2 or 3
Stm12	If the number of days between the recipient's interview and birthday (inclusive) is between 0 and 365, child drug recency must not equal 3
Stm13	If recipient's age at first parent drug use equals his or her current age or the recipient's age at first child drug use equals his or her current age or the recipient's number of days between his or her interview date and date at first child drug use less than 30, then donor's recency must not equal 3
Stm14	Donor must be a past month or past year (but not past month) child drug user (child drug recency = 1 or 2)
Stm15	Donor's proportion of past year child drug use * recipient's max number of days could have used child drug in past year must be less than (or equal to) the recipient's maximum possible past year child drug frequency of use. The 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 the recipient could have used child drug, as determined by the month of first use (2) if the 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)
Stm16	Donor's proportion of past year child drug use * recipient's min number of days could have used child drug in past year must be greater than (or equal to) the 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.)
Stm17	(Recipient's proportion of past year child drug use * max number of days could have used child drug in past year) must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
Stm18	If recipient's age at first child drug use equals his or her current age, (1) recipient's donor's proportion of past year child drug use * recipient's max number of days could have used child drug in 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 child drug use * recipient's max number of days could have used child drug in past year cannot be greater than the recipient's days between his or her interview date and birthday (inclusive)

Table H.19 Constraints for Stimulants, Methamphetamines, Pain Relievers, and OxyContin (continued)

Constraint #	Constraint
Stm19	If recipient's age at first child drug use equals his or her current age, (1) recipient's donor's proportion of past year child drug use* recipient's max number of days could have used child drug in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (minus 29) and (2) donor's proportion of past year child drug use * recipient's max number of days could have used child drug in past year cannot be greater than the recipient's days between his or her interview date and birthday (minus 29)
Stm20	Donor must be a past month child drug user (child drug recency = 1)
Stm21	Donor must be a past year (but not past month) child drug user (child drug recency = 2)
Stm22	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)
Stm23	If recipient's month/year of first use data for the parent drug indicate that he/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 max number of days could have used the parent drug in past year must be greater than 1
Stm24	If recipient's month/year of first use data for the child drug indicate that he/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 max number of days could have used child drug in past year must be greater than 1
Stm25	If recipient's month/year of first use data for the parent drug indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month parent drug user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than 1
Stm26	If recipient's month/year of first use data for the child drug indicate that he/she must have used at least once in the interval (1 year before interview, 30 days before interview), then if donor is a past month child drug user, donor's proportion of past year use * recipient's max number of days could have used in past year must be greater than 1
Stm27	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
Stm28	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

Table H.20 Restrictions and Portion of the Predictive Mean Vector for Stimulant and Pain Reliever Users (Including Methamphetamines and OxyContin)

Missingness Pattern					Number of Cases	Constraints	Predictive Mean Vector[1]
#	Stimulants or Pain Reliever Recency	Meth. or Oxy. Recency	Stimulants or Pain Reliever 12-Month Freq.	Meth. or Oxy. 12-Month Freq.			
1	(Past month)		Missing		Pain Relievers: 71 Stimulants: 11	(Stm1-5,23)	1. PY
2	(Past year but not past month)		Missing		Pain Relievers: 50 Stimulants: 6	(Stm1-3,6-7)	1. PY
3	Past year				Pain Relievers: 5 Stimulants: 3	(Stm7-8)	1. R1/(R1+R2)
4	Past year		Missing		Pain Relievers: 23 Stimulants: 6	(Stm1-3,5,7-8,25)	1. R1/(R1+R2) 2. PY
5	Missing (lifetime use known)		Missing		Pain Relievers: 314 Stimulants: 72	(Stm1-3,5,7,25)	1. R1 2. R2 3. (R1+R2)*PY
5	Missing (lifetime use imputed)		Missing		Pain Relievers: 10 Stimulants: 2		
6	(Past month)	(Past month)		Missing	Pain Relievers: 0 Stimulants: 2	(Stm4,15-20,24)	PY
7	(Past year not missing)	(Past year not past month)		Missing	Pain Relievers: 1 Stimulants: 1	(Stm8,14-20)	PY
8	(Past year not missing)	Past year			Pain Relievers: 0 Stimulants: 0	(Stm7-8)	1. R1/(R1+R2)
9	(Past year not missing)	Past year	Missing		Pain Relievers: 0 Stimulants: 0	(Stm1-3,7-8,23)	1. R1/(R1+R2) 2. PY
10	(Past year not missing)	Past year		Missing	Pain Relievers: 7 Stimulants: 0	(Stm7-8,15-17,26)	1. R1/(R1+R2) 2. PY

Table H.20 Restrictions and Portion of the Predictive Mean Vector for Stimulant and Pain Reliever Users (Including Methamphetamines and OxyContin) (continued)

Missingness Pattern							
#	Stimulants or Pain Reliever Recency	Meth. or Oxy. Recency	Stimulants or Pain Reliever 12-Month Freq.	Meth. or Oxy. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
11	(Past year not missing)	Past year	Missing	Missing	Pain Relievers: 0 Stimulants: 1	(Stm1-3,7-8,15-17,23,28)	1. R1/(R1+R2) 2. PY
12	(Past year not missing)	Missing (lifetime use known)		Missing	Pain Relievers: 14 Stimulants: 0	(Stm7-8,15-17)	1. R1 2. R2 3. (R1+R2)*PY
12	(Past year not missing)	Missing (lifetime use imputed)		Missing	Pain Relievers: 1 Stimulants: 1		
13	(Past month)	(Past month)	Missing	Missing	Pain Relievers: 1 Stimulants: 1	(Stm1-4,7,15-17,20,23-24)	PY
14	(Past month)	(Past year not past month)	Missing	Missing	Pain Relievers: 0 Stimulants: 0	(Stm1-4,7,15-17,21,23)	PY
15	(Past year not past month)	(Past year not past month)	Missing	Missing	Pain Relievers: 2 Stimulants: 3	(Stm1-3,6-7,15-17,21)	PY
16	Past year	Past year			Pain Relievers: 3	(Stm7-8,14,19,28)	R1/(R1+R2)
17	Past year	Past year	Missing		Pain Relievers: 0 Stimulants: 0	(Stm1-3,7-8,14,19,25,28)	1. R1/(R1+R2) 2. PY

Table H.20 Restrictions and Portion of the Predictive Mean Vector for Stimulant and Pain Reliever Users (Including Methamphetamines and OxyContin) (continued)

Missingness Pattern							
#	Stimulants or Pain Reliever Recency	Meth. Or Oxy. Recency	Stimulants or Pain Reliever 12-Month Freq.	Meth. or Oxy. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
18	Past year	Past year		Missing	Pain Relievers: 7 Stimulants: 2	(Stm7-8,14-17,19,26,28)	1. R1/(R1+R2) 2. PY
19	Past year	Past year	Missing	Missing	Pain Relievers: 3 Stimulants: 2	(Stm1-3,7-8,14-17,19,25-26,28)	1. R1/(R1+R2) 2. PY
20	Past year	Missing (lifetime use known)		Missing	Pain Relievers: 16 Stimulants: 1	(Stm7-8,10,15-17,19,28)	1. R1/(R1+R2) 2. PY
20	Past year	Missing (lifetime use imputed)		Missing	Pain Relievers: 0 Stimulants: 0		
21	Past year	Missing (lifetime use known)	Missing	Missing	Pain Relievers: 1 Stimulants: 1	(Stm1-3,7-8,10,15-17,19,25,28)	1. R1/(R1+R2) 2. PY
21	Past year	Missing (lifetime use imputed)	Missing	Missing	Pain Relievers: 0 Stimulants: 0		

Table H.20 Restrictions and Portion of the Predictive Mean Vector for Stimulant and Pain Reliever Users (Including Methamphetamines and OxyContin) (continued)

Missingness Pattern							
#	Stimulants or Pain Reliever Recency	Meth. Or Oxy. Recency	Stimulants or Pain Reliever 12-Month Freq.	Meth. or Oxy. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
22	(Past month)	Missing (lifetime use known)		Missing	Pain Relievers: 0	(Stm4,7,10,15-17,19)	1. R1 2. R2 3. (R1+R2)*PY
					Stimulants: 0		
22	(Past month)	Missing (lifetime use imputed)		Missing	Pain Relievers: 0		
					Stimulants: 0		
23	(Past month)	Missing (lifetime use known)	Missing	Missing	Pain Relievers: 1	(Stm1-4,7,10,15-17,19,23)	1. R1 2. R2 3. (R1+R2)*PY
					Stimulants: 0		
23	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Pain Relievers: 1		
					Stimulants: 0		
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Pain Relievers: 29	(Stm1-3,7,10,15-17,19,22,28)	1. R1 2. R2 3. (R1+R2)*PY
					Stimulants: 36		
24	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	Pain Relievers: 1		
					Stimulants: 0		

Table H.20 Restrictions and Portion of the Predictive Mean Vector for Stimulant and Pain Reliever Users (Including Methamphetamines and OxyContin) (continued)

Missingness Pattern							
#	Stimulants or Pain Reliever Recency	Meth. Or Oxy. Recency	Stimulants or Pain Reliever 12-Month Freq.	Meth. or Oxy. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
24	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	Pain Relievers: 0 Stimulants: 0	(Stm1-3,7,10,15-17,19,22,28)	1. R1 2. R2 3. (R1+R2)*PY
24	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	Pain Relievers: 0 Stimulants: 4		
25	Past year	(Past year not past month)	Missing	Missing	Pain Relievers: 0 Stimulants: 2	(Stm1-3,7-8,14-17,19,25)	1. R1/(R1+R2) 2. PY
	Lifetime user, nothing missing				Pain Relievers: 10,657 Stimulants: 5,213		
	Imputed to lifetime nonuse				Pain Relievers: 241 Stimulants: 134		
	Lifetime nonuser, nothing missing				Pain Relievers: 56,849 Stimulants: 62,804		

NOTE: Users of stimulants included users of methamphetamines.

[1] The predictive mean vector components are defined by the following:

1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

H.2.5 Source of Income

There was a large number of missingness patterns for the source of income variables because they were imputed simultaneously in a set. The only logical constraint applied to the potential donors was that they have the same value as the recipient for the imputation-revised family skip variable (IRFAMSKP). This logical constraint was applied for all missingness patterns.

Table H.21 Restrictions and Portion of the Predictive Mean Vector for Income

Missingness Pattern				Number of Cases	Constraints	Predictive Mean Vector[1]
#	Welfare Months	Family Payment	Family Service			
1	Missing	Receiving	Not Receiving	189	IRFAMSKP of donor equal to that of recipient	WMS, and probabilities associated with other missing elements
2	Missing	Not Receiving	Receiving			
3	Missing	Receiving	Receiving			
4	Missing	Not Receiving	Missing	127		SVC*WMS, SVC, and probabilities associated with other missing elements
5	Missing	Missing	Not Receiving	207		PMT*WMS, PMT, and probabilities associated with other missing elements
6	Missing	Missing	Missing	353		[1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements

[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)

H.2.6 Health Insurance

Both of the methods that were used to create the final imputation-revised health insurance variables, the "Old Method" and the "Constituent Variables Method," are described in this section (see Chapter 10 for details).

H.2.6.1 Health Insurance (Old Method)

The health insurance variables IRINSUR (overall health insurance using only questions available in the 1999 questionnaire), IRINSUR3 (overall health insurance using all questions available in the 2001 questionnaire and beyond), and IRPINSUR (private health insurance) were imputed as a set. Their edited counterparts were INSUR, INSUR3, and PINSUR. Details are in Chapter 10.

Table H.22 Constraints for Health Insurance (Old Method)

Constraint #	Logical Constraint
HI2001_1	Donor must not have received private health insurance (PINSUR = 0)[1]
HI2001_2	Donor must not have received overall health insurance by the 1999 definition (INSUR = 0)
HI2001_3	Donor must have received overall health insurance by the 2001 definition (INSUR3 = 1)
HI2001_4	Donor must have received overall health insurance by the 1999 definition (INSUR = 1)[1]

[1]Technically, these were not logical constraints. See Chapter 7 for details.

Table H.23 Health Insurance (Old Method)

Missingness Pattern				Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	INSUR3	INSUR	PINSUR			
1	Missing	No	No	75	HI2001_1, HI2001_2	(OVR*(1-PRV))/(1-OVR*PRV)
2	Yes	Missing	No	24	HI2001_1, HI2001_3	(OVR*(1-PRV))/(1-OVR*PRV)
3	Missing	Missing	No	100	HI2001_1	(OVR*(1-PRV))/(1-OVR*PRV)
4	Yes	Missing	Missing	14	HI2001_3	OVR, OVR*PRV
5	Missing	Missing	Missing	338		OVR, OVR*PRV
6	Yes	Yes	Missing	68	HI2001_4	PRV

[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)

H.2.6.2 Health Insurance (Constituent Variables Method)

The health insurance variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT were imputed as a set. Their edited counterparts were CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Details are provided in Chapter 10. The "Predictive Mean Vector" column is omitted from Table H.24 because the elements of the vector were simply the predictive means associated with all missing variables. For example, for all missingness patterns where CAIDCHIP was missing, the probability that the respondent had CAIDCHIP = 1 was included in the predictive mean vector. The "Logical Constraints" column also is omitted from Table H.24 because no logical constraints were applied.

Table H.24 Health Insurance (Constituent Variables Method)

Missingness Pattern					Number of Cases
#	CAIDCHIP	MEDICARE	CHAMPUS	PRVHLTIN	
1	Missing				227
2		Missing			62
3	Missing	Missing			36
4			Missing		57
5	Missing		Missing		24
6		Missing	Missing		1
7	Missing	Missing	Missing		14
8				Missing	193
9	Missing			Missing	59
10		Missing		Missing	11
11	Missing	Missing		Missing	19
12			Missing	Missing	18
13	Missing		Missing	Missing	48
14		Missing	Missing	Missing	3
15	Missing	Missing	Missing	Missing	69

Appendix I: Quality Control Measures Used in the Imputation Procedures

Appendix I: Quality Control Measures Used in the Imputation Procedures

I.1 Introduction

For the 2005 National Survey of Drug Use and Health (NSDUH),[177] the quality control (QC) imputation procedures as applied to demographic, drug use, income, health insurance, nicotine dependence, and household composition (roster) variables are discussed in this appendix. The imputation process involved three basic procedures: (1) weight adjustment for item nonresponse in the models, (2) predictive mean modeling, and (3) final assignment of imputed values using these predicted means. Drug use variables had an additional step to randomly assign the date of first drug use. QC measures were performed at each of these steps. Besides these QC measures, specific checklists for demographic, drug use, income, health insurance, nicotine dependence, and roster variables were used during the imputation procedures in the 2005 survey. These checklists provided formal documentation of the QC checks that were implemented during imputation.

In addition to the QC checks listed below, all SAS[178] programs, which were run by members of the imputation team, were subsequently reviewed for errors by the person who ran the SAS programs and an independent reviewer. Messages in the SAS log file, model convergence, and missing values were some of the noticeable problems that were examined. The imputation team also edited demographic variables (age, interview date, birth date, gender, race, and Hispanicity) and household composition variables. QC measures were implemented throughout these editing processes, and specific checklists were developed for the editing of demographics and roster variables. However, the QC procedures that were used in the editing process will not be discussed in this appendix.[179] The imputation team performed QC checks when delivering variables to other NSDUH teams. Checklists were developed for the regression imputation method used for cigarette dependency variables. These checks for delivering variables and the regression imputation method will also not be discussed in this chapter. Yet, checks involved in each of the predictive mean neighborhood (PMN) imputation steps and the random assignment of drug age of first use are described in detail in the following sections.

I.2 Step 1. Weight Adjustment for Item Nonresponse in Models

In this step, it was necessary to define a set of variables where item nonresponse was characterized. To have been classified as a "complete" respondent, a person would have had to respond to all the questions within the variable set. Only complete respondents were used to build the models in the next step. As a general practice, the weights were adjusted so that the weights for complete respondents represented the entire domain, where "domain" was defined as

---

[177] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

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

[179] See the logical editing procedures used to create these variables in Chapters 4 and 8 of this report. For more details on other editing procedures that were performed on NSDUH data prior to imputation, see Kroutil, Handley, Suresh, Felts, and Bradshaw (2007).

the population of interest (e.g., lifetime users aged 12 to 17 years old). This was accomplished by using an item response propensity model, a special case of the generalized exponential model (GEM),[180] which is described in greater detail in Appendix B. For this step, QC measures were conducted as follows:

- 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 model 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, for example to see whether the base variables were present if interactions existed. For example, if the variable representing age was dropped, then the interaction between age and gender also would have been dropped.
- 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 weights. If the maximum adjustment was too high (usually greater than 3), 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 so that any final adjustment applied was acceptable.
- After the weights were adjusted, the ratio of the maximum adjusted weight to the mean adjusted weight (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.
- Unequal weighting effect (UWE) was checked before and after adjustment to ensure there was no significant variance increase due to the nonresponse adjustment. The difference in the UWE after adjustment value should have been 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 people identified as item nonrespondents was recorded. This number was checked to ensure that it was the same as the number of people who were excluded from the model-building process.
- When using PROC MEANS, 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 other members of the imputation team.

---

[180] The GEM macro, which was written in SAS/IML® software, was developed at RTI International (a trade name of Research Triangle Institute) for weighting procedures.

I.3 Step 2. Predictive Mean Modeling

For each question that used the PMN imputation method, modeling procedures were used to determine the predicted 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 gate question.[181] Although only item respondents contributed to the model, predicted mean values were determined regardless of whether the respondent 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 the main body of this report. For this predictive mean modeling step, the following QC measures were employed:

- 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.[182]
- All models were checked for singularities and collinearities. Any singularities that occurred were investigated and the model 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.[183] If there was a "data warning" message in the log, the SUDAAN model was 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 were also 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 have exceeded 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.

[181] 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.

[182] Although the CLASS statement could have been used in SAS® to automatically create the appropriate indicator variables, no such option was available in SAS®-callable SUDAAN® (Release 9.0), which was used to fit the polytomous logistic regression models.

[183] Greater details can be found in the *SUDAAN User's Manual: Release 9.0* (RTI, 2004).

- 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 other members of the imputation team.

I.4 Step 3. Final Assignment of Imputed Values

I.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 needed to have very similar predicted means. QC checks in this step had two objectives: (1) to ensure that the imputed values were consistent with preexisting nonmissing values and (2) to ensure that the imputed values were assigned as intended. The following checks 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 to revise the programs so that a donor could have been 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 within 1 percent 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.[184]

- The frequency of the variable "WORKED" was checked to ensure that no values were equal to zero. Values greater than zero 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 had been implemented, the distribution of values for nonrespondents was checked against the distribution of values for all respondents to ensure the similarity of these two items.

[184] Refer to Appendix G for more details about likeness restrictions and the "SMALLFLG" variable.

- 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 had a nonapplicable 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 other members of the imputation team.

I.4.2 Specific Imputation Checks for Univariate Predictive Mean Neighborhood

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 use drug use, finer income variables, questionnaire roster variables, and "Constituent Variables Method" for imputing health insurance variables. The UPMN utilized 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 output was checked for the items provided in the following list:

- The imputed values were checked against preexisting nonmissing values for consistency. Listed below are a few checks 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[185] drugs, the imputed variables of the parent drug were crossed with those of the child drug(s) to ensure consistency.
  - For parent-child drugs, the parent drug's age at first use must have been less than or equal to the child drug's age at first use.
  - The respondent's age at first drug use must not have equaled the respondent's age, if the recency was "not in the past year."
  - The imputed number of people in household younger than age 18 should have been 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.

[185] 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.

  - 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 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, edited number of people in household aged 65 or older (HH65) was compared with imputed number of people in household aged 65 or older (IRHH65) to ensure that IRHH65 had no missing values.

I.4.3 Specific Imputation Checks for Multivariate Predictive Mean Neighborhoods

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, and recency and frequency of drug use. For these multivariate imputations, the items provided in the following list 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. Those 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 have equaled or exceeded the 30-day frequency.
  - Past month users must have had a valid 30-day frequency (not a skip code).
  - Past year users must have had a valid 12-month frequency (not a skip code).
  - For alcohol, 30-day frequency must have exceeded or equaled the "binge" drinking frequency.
  - For parent-child drugs (e.g., cocaine and crack; smokeless tobacco and snuff), the parent drug recency must have occurred no later than the child drug's recency.
  - For cocaine and crack, the cocaine 12-month frequency must have equaled or exceeded the crack 12-month frequency, if it existed.
  - For cocaine and crack, the cocaine 30-day frequency must have equaled or exceeded the crack 30-day frequency, if it existed.
  - The recency and frequency of use variables that were imputed must have been 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 have exceeded the maximum usage period less 30.

- If the respondent's age was equal to the age at first use, the recency of use must have been imputed to be past month or past year not past month.

- In the questionnaire, 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 have been 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 have been consistent with the time period between the birthday and the interview date, and it must have been consistent with the month that the respondent began using, if available.

- For income, only people 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 have been 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 variable employment status (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 have been 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.

I.5 Additional Step for Drug Variables: Assignment of the Date of First Drug Use

For the age at first drug use imputations, an additional step was required that assigned a date of first use. QC checks in this step had two objectives: (1) the assigned date must have been consistent with the imputed age at first use, and (2) the assigned date must have been consistent with other imputation-revised drug variables, such as recency and frequency variables.

- The assigned date of first use should have been consistent with the given birth date and the imputation-revised age at first use.
- The assigned date of first use should have been consistent with the given interview date and the imputation-revised recency/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 have occurred if the birthday or interview date occurred on the 1st 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 year/months were being transmitted to the imputation-revised year/month of first use.
- A frequency of the imputation-revised month/day/year of first use variables was run to ensure that all were within the acceptable numbers (i.e., month was between 1 and 12, or 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 the consistency.

Sometimes, because an error was discovered further along in the process, a patch was necessary for earlier imputations. When the variables were reimputed and the dataset was updated, it was crucial to compare the old (incorrect) imputation-revised variable and the new corrected variable with the reimputed values. This was necessary to ensure that (1) the changes made were within expected limits and that (2) other cases did not inadvertently change with the correction. Cases that had 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 other members of the imputation team.

I.6 Imputation Checklists

Most of the QC measures above mentioned were incorporated into specific imputation checklists for demographic, drug use, income, health insurance, nicotine dependence, and roster variables. These checklists included a technician check, where the person 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, another person (reviewer) performed an independent check of the same item. 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 checks that were incorporated. Checklists also were updated and revised to reflect the changes in the programs before each processing cycle. Furthermore, 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 in the previous surveys. For the 2005 survey, additional checklists were constructed; thus, almost all major imputation programs and, as a result, all variable categories were covered. The specific checklists that were implemented for 2005 NSDUH imputation programs are summarized in Table I.1.

Table I.1 Summaries of Checklists for Imputation Programs for the 2005 NSDUH

PROGRAMS	**SUBTASKS**					
	Demographics	**Drug**	**Income**	**Health Insurance**	**Roster**	**Nicotine Dependence**
Editing	Core demographics editing[1]	N/A[2]	N/A[2]	N/A[2]	Roster Editing[1]	N/A[2]
Item nonresponse weight adjustment	√	√	√	√	√	Regression*
Predictive mean modeling	√	√	√	√	√	
UPMN	√	√	√	√	√	
MPMN	√	√	√	√	N/A[2]	
Date of first drug use	N/A[2]	√	N/A[2]	N/A[2]	N/A[2]	N/A[2]
Delivering variable	√	√	√	√	√	√

NOTE: "√" implies that a specific checklist was developed for this subtask (column header) and program category (row header).
N/A = not applicable.
[1] Specific editing checklists were developed for the core demographics and roster editing, and regression imputation method of nicotine dependence variables.
[2] Program category (row header) is not applicable for this subtask (column header).

Appendix J: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

Appendix J: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

J.1 Introduction

In the household roster for the 2005 National Survey on Drug Use and Health (NSDUH),[186] 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 did not make sense 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. 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. However, beginning with the 2000 survey and in every survey year since then, consistency checks have been added to the CAI instrument that allowed the interviewer, if needed, to correct the error while giving the interview. Details about these consistency checks are presented in Chapter 8 of the main body of this report.

In general, two types of consistency checks were implemented in the 2005 survey. The first type compared the entry in the roster with previously entered questionnaire information, specifically the respondent's age (CURNTAGE) and gender, and the second type checked for internal consistency within the household roster. In some cases, a consistency check would have been triggered even though the response was legitimate. This occurred if CURNTAGE was considered incorrect, or it occurred in extremely rare family situations such as a stepmother who was younger than her stepson. With the exception of the check against the previously entered respondent's gender, the interviewer could have overridden the consistency check and explained 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 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.

J.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

[186] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

of age that was stored by Blaise.[187] Explanations given by interviewers for overrides to consistency checks against CURNTAGE are provided in Table J.1. Because CURNTAGE 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 the final questionnaire-edited age (AGE), the derivation of which is described in Chapter 4 of the main body of this report.

In the 2002 survey, the explanations provided in Table J.1 were not reviewed when determining AGE, nor were they reviewed when determining the final value for the age of the "self" entry in the roster. However, since the 2003 survey, these explanations have been carefully reviewed. In rare cases, the final value for age (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 how this was done are in Chapter 4.

Even in cases where the explanation seemed clear that CURNTAGE was wrong, 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, since some of the differences could be due to the fact that a birthday could have occurred between the drug modules and the roster.[188] In other situations, the value of CURNTAGE was wrong, 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 "Comments" column in bolded italics. Otherwise, the reason for not changing the value of AGE to roster age also is shown in this column. The last column in Table J.1 indicates whether the roster of the other pair member, if it existed, supported CURNTAGE or the override age as the respondent's age.

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

[188] 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 J.1 Explanations for Overrides to Consistency Checks against CURNTAGE

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE = Final Roster Age	Verbatim Explanation from Field Interviewers[1]	Comment[2]	Respondent's Age in Roster of Other Pair Member
1	26	25	25	26	25 is correct	Diff. ≤ 1 year	Not in a pair
2	22	21	22	22	Subject provided incorrect age on last answer.	Diff. ≤ 1 year	22
3	25	23	25	25	HE is the respondent I am talking to.	Diff. > 1 year; other pair member supports NEW AGE	25
4	14	13	13	14	Father made a mistake on daughter's age	Diff. ≤ 1 year	14
5	33	34	34	33	age was not correct orginally it is in fact 34	Diff. ≤ 1 year	33
6	18	17	17	18	rsp is 18	Diff. ≤ 1 year	18
7	39	38	39	39	wrong age in roster	Diff. ≤ 1 year	Not in a pair
8	33	25	26	26	PERSON IS ILLITERATE, SHOWED ME HIS ID, BUT SR HAD HIM AS BEING 26 YR. OLD	Diff. > 1 year; screener & FI support NEW AGE	Not in a pair
9	40	41	41	40	he is actually 41 yrs old	Diff. ≤ 1 year	40
10	30	31	31	30	r said she is 30 not 31	Diff. ≤ 1 year	30
11	13	12	13	13	resp is 12	Diff. ≤ 1 year	Not in a pair
12	28	27	27	28	R gave FI wrong answer	Diff. ≤ 1 year	Not in a pair
13	20	19	20	20	resp was 19 and is now 20	Diff. ≤ 1 year	20
14	52	51	51	52	he is 52	Diff. ≤ 1 year	Not in a pair
15	31	30	31	31	30	Diff. ≤ 1 year	31
16	21	20	20	21	she said she was 21 and then now says she is 20	Diff. ≤ 1 year	21
17	26	32	25	26	touched wrong tabs	Diff. > 1 year; screener age supports NEW AGE	Not in a pair

Table J.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE = Final Roster Age	Verbatim Explanation from Field Interviewers[1]	Comment[2]	Respondent's Age in Roster of Other Pair Member
18	34	37	36	37	her age is 37	***AGE was changed to equal roster age***	Not in a pair
19	20	21	21	20	both he and he's 21	Diff. ≤ 1 year	20
20	51	52	52	51	Because error was made at beginning of interview respondent is 52	Diff. ≤ 1 year	51
21	40	41	40	40	entered 40 computer confirmed and now says 41 was entered	Diff. ≤ 1 year	40
22	13	12	13	13	roster has been confused	Diff. ≤ 1 year	13
23	43	42	42	43	r say sept comm she will be 43	Diff. ≤ 1 year	43
24	28	29	28	28	screening resp forgot exact ages	Diff. ≤ 1 year	28
25	35	36	36	35	husband gave wrong age	Diff. ≤ 1 year	35
26	24	20	19	20	no need for this pop-up box	Diff. ≤ 1 year	20
27	26	25	25	26	R BECOME 26 ON MONDAY 6-6-05	Diff. ≤ 1 year	Not in a pair
28	45	46	46	45	46 is correct	Diff. ≤ 1 year	Not in a pair
29	72	66	72	72	seem to grt the ages correctly entered into the computer 72, 66, 49,41	Diff. > 1 year; screener age supports NEW AGE	Not in a pair
30	24	23	23	24	i was trying to follow directions & make the age 24 consistent	Diff. ≤ 1 year	Not in a pair
31	17	18	18	17	r's brother told to me the r was 18	Diff. ≤ 1 year	17
32	49	48	48	49	she made error	Diff. ≤ 1 year	49

Table J.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE = Final Roster Age	Verbatim Explanation from Field Interviewers[1]	Comment[2]	Respondent's Age in Roster of Other Pair Member
33	51	52	52	51	51	Diff. ≤ 1 year	51
34	48	47	47	48	she said she was 47	Diff. ≤ 1 year	48
35	27	28	28	27	confirmed age earlier and now confirms 28	Diff. ≤ 1 year	Not in a pair
36	33	32	32	33	last age is correct	Diff. ≤ 1 year	Not in a pair
37	17	18	18	17	sr roster respondent as 18 but respondent ind she is 17.S	Diff. ≤ 1 year	17
38	22	21	22	22	age is 22	Diff. ≤ 1 year	22
39	79	80	80	79	r stated he was 80	Diff. ≤ 1 year	79
40	17	16	16	17	person turned 17 after screening	Diff. ≤ 1 year	17
41	68	69	69	68	error	Diff. ≤ 1 year	Not in a pair
42	13	20	13	13	My Mistake	Diff. > 1 year; screener & FI support NEW AGE	Not in a pair
43	19	18	19	19	i could not get to roster-the correct age is 18	Diff. ≤ 1 year	19
44	33	25	25	25	respondant is 25 and female	***AGE was changed to equal roster age***	Not in a pair
45	19	18	19	19	given wrong answer	Diff. ≤ 1 year	Not in a pair
46	14	13	14	14	the R is 13 they are twins, don't know where 14 came from	Diff. ≤ 1 year	14

[1] These entries came directly from the 2005 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and 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.

J.3 Override Comments from Interviewers: Internal Consistency Check Overrides

Internal consistency checks in the roster were performed. Explanations by interviewers for overrides to internal consistency checks are provided in Table J.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 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.

Table J.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	20	48-year-old grandparent	step grandfather	Legitimate; interviewer's override stands
2	Respondent has multiple spouses or live-in partners	33	67-year-old spouse	both husband and boyfriend live with R	Legitimate; interviewer's override stands
3	Respondent has multiple spouses or live-in partners	33	29-year-old live-in partner	both husband and boyfriend live with R	Legitimate; interviewer's override stands
4	Respondent is 16 or younger & has an in-law	16	44-year-old parent-in-law	married 16yr old	Legitimate; interviewer's override stands
5	Respondent is 16 or younger & married or has a live-in partner	16	22-year-old spouse	married to 16 yr old	Legitimate; interviewer's override stands
6	Respondent is 16 or younger & married or has a live-in partner	16	22-year-old live-in partner	teen parent living on own w/child and child's father	Legitimate; interviewer's override stands
7	Grandparent and respondent less than 30 years apart	14	31-year-old grandparent	31 male is step grandfather	Legitimate; interviewer's override stands
8	Respondent is 16 or younger & married or has a live-in partner	16	19-year-old live-in partner	live in boy friend	Legitimate; interviewer's override stands
9	Respondent is 16 or younger & married or has a live-in partner	16	19-year-old spouse	R is married to the 19 year old house member	Legitimate; interviewer's override stands
10	Respondent's wife is 16 or younger	17	15-year-old live-in partner	mother allows this relationship	Legitimate; interviewer's override stands
11	Respondent's wife is 16 or younger	22	16-year-old live-in partner	immancipated minor	Legitimate; interviewer's override stands
12	Respondent is 16 or younger & married or has a live-in partner	16	20-year-old live-in partner	R IS PREGNANT WITH HIS CHILD	Legitimate; interviewer's override stands
13	Grandparent and respondent less than 30 years apart	16	40-year-old grandparent	STEP GRANDFATHER	Legitimate; interviewer's override stands

Table J.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
14	Respondent's female in-law is 16 or younger	44	16-year-old daughter-in-law	MARRIED TO SON, EMANCIPTED MINOR	Legitimate; interviewer's override stands
15	Respondent's wife is 16 or younger	22	16-year-old spouse	verified by husband	Legitimate; interviewer's override stands
16	Respondent is 16 or younger & married or has a live-in partner	15	17-year-old live-in partner	mother allows this relationship	Legitimate; interviewer's override stands
17	Respondent's wife is 16 or younger	19	16-year-old spouse	R is married to the 16 year old household member	Legitimate; interviewer's override stands
18	Respondent is 16 or younger & married or has a live-in partner	16	22-year-old spouse	got married at a young age	Legitimate; interviewer's override stands
19	Respondent's wife is 16 or younger	20	16-year-old spouse	wife is actually 16 years old	Legitimate; interviewer's override stands
20	Respondent is 16 or younger & married or has a live-in partner	16	20-year-old spouse	16 year old is his wife	Legitimate; interviewer's override stands
21	Grandparent/Grandchild and R less than 30 years apart	14	40-year-old grandparent	he is step grandfather	Legitimate; interviewer's override stands
22	Respondent's wife is 16 or younger & Respondent is 16 or younger & married or has a live-in partner	15	15-year-old live-in partner	lives with 15 yr old girlfriend at her mothers house	Legitimate; interviewer's override stands
23	Grandparent and respondent less than 30 years apart	16	42-year-old grandparent	This is his step-grandfather	Legitimate; interviewer's override stands
24	Respondent has multiple spouses or live-in partners	44	50-year-old live-in partner	this is what R has told me	Legitimate; interviewer's override stands
25	Respondent has multiple spouses or live-in partners	44	32-year-old live-in partner	this is what R has told me	Legitimate; interviewer's override stands
26	Grandparent and respondent less than 30 years apart	13	42-year-old grandparent	Info correct	Legitimate; interviewer's override stands
27	Grandparent and respondent less than 30 years apart	16	45-year-old grandparent	Grandma & daughter teen pregancy	Legitimate; interviewer's override stands
28	Respondent's daughter is less than 13 years younger than respondent	13	3-month-old child	respondent became pregnant age 12	Legitimate; interviewer's override stands

Table J.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
29	Respondent's daughter is less than 13 years younger than respondent	15	5-month-old child[2]	R has a 5 month old daughter	Legitimate; interviewer's override stands
30	Respondent's wife is 16 or younger	20	16-year-old spouse	married to person 1	Legitimate; interviewer's override stands
31	Grandparent and respondent less than 30 years apart	24	48-year-old grandparent	not biological grandfather	Legitimate; interviewer's override stands
32	Grandparent and respondent less than 30 years apart	13	35-year-old grandparent	she is the step-grandmother and considered grandma	Legitimate; interviewer's override stands
33	Respondent is 16 or younger & married or has a live-in partner	16	18-year-old live-in partner	they live together	Legitimate; interviewer's override stands
34	Respondent's husband is 16 or younger	18	16-year-old live-in partner	this is her live in boyfriend	Legitimate; interviewer's override stands
35	Gap of 25 or more years between sister and respondent	13	38-year-old sibling	she is his biological sister and that is her age	Legitimate; interviewer's override stands
36	Respondent is 16 or younger & married or has a live-in partner	16	20-year-old live-in partner	R IS AN EMANCIPATED MINOR LIVING WITH BOYFRIEND & HIS FATHER	Legitimate; interviewer's override stands
37	Respondent's husband is 16 or younger	23	16-year-old live-in partner	Because the 23 yr old is living with a 16 yr. old. according to the R.	Legitimate; interviewer's override stands
38	Respondent's son is less than 13 years younger than respondent	13	3-month-old child	she is 13 and has a son 3 mths old	Legitimate; interviewer's override stands
39	Respondent is 16 or younger & married or has a live-in partner	14	40-year-old live-in partner	They are not married	Overrule; probable parent
40	Respondent's female in-law is 16 or younger	22	3-year-old child-in-law	Child is respondents daughter-in-law	Overrule; unsure of relationship (85)
41	Respondent is 16 or younger & married or has a live-in partner	16	21-year-old spouse	interviewee is married	Legitimate; interviewer's override stands
42	Respondent has multiple spouses or live-in partners	35	51-year-old spouse	This is a 3 way intimate relationship	Legitimate; interviewer's override stands

Table J.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
43	Respondent has multiple spouses or live-in partners	35	43-year-old live-in partner	This is a 3 way intimate relationship	Legitimate; interviewer's override stands
44	Grandparent and respondent less than 30 years apart	15	26-year-old grandparent	grandfather's wife is 26 years old and therefore is grandmother to this R	Legitimate; interviewer's override stands
45	Respondent is 16 or younger & married or has a live-in partner	16	21-year-old live-in partner	16 year old r is living with 21 year old male and have a 5 month old son with plans to marry	Legitimate; interviewer's override stands
46	Gap of 25 or more years between sister and respondent	12	38-year-old sibling	mom verifies this	Legitimate; interviewer's override stands
47	Gap of 25 or more years between sister and respondent	17	42-year-old sibling	she said the 42 year old is her full sister	Legitimate; interviewer's override stands
48	Respondent has multiple spouses or live-in partners	44	60-year-old live-in partner	Tri-Relatioship	Legitimate; interviewer's override stands
49	Respondent has multiple spouses or live-in partners	44	52-year-old live-in partner	Tri-Relatioship	Legitimate; interviewer's override stands
50	Grandchild and respondent less than 30 years apart	29	4-month old grandchild	r said this is her 14 yrs old daughts son	Legitimate; interviewer's override stands
51	Grandparent and respondent less than 30 years apart	20	42-year-old grandparent	this is the age I was told	Legitimate; interviewer's override stands
52	Respondent's daughter is less than 13 years younger than respondent	21	14-year-old child	14	Overrule; probable sibling
53	Respondent is 16 or younger & married or has a live-in partner	16	19-year-old live-in partner	She is living here and is engaged to the 19 year old male	Legitimate; interviewer's override stands
54	Respondent is 16 or younger & married or has a live-in partner	16	19-year-old live-in partner	respondants are living partners	Legitimate; interviewer's override stands
55	Respondent's wife is 16 or younger	19	16-year-old live-in partner	16 year old is imamcipated minor they blive together as partners	Legitimate; interviewer's override stands
56	Respondent has multiple spouses or live-in partners	23	23-year-old live-in partner	Religion Islaam	Legitimate; interviewer's override stands

Table J.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
57	Respondent has multiple spouses or live-in partners	23	20-year-old live-in partner	Religion Islaam	Legitimate; interviewer's override stands
58	Grandparent and respondent less than 30 years apart	18	44-year-old grandparent	This is a step grandparent	Legitimate; interviewer's override stands
59	Respondent has multiple spouses or live-in partners	36	36-year-old spouse	r states that this is accurate	Legitimate; interviewer's override stands
60	Respondent has multiple spouses or live-in partners	36	29-year-old live-in partner	r states that this is accurate	Legitimate; interviewer's override stands
61	Respondent's wife is 16 or younger	17	16-year-old live-in partner	have a baby together,living w/girls mother until own home is ready	Legitimate; interviewer's override stands
62	Respondent's husband is 16 or younger	17	16-year-old live-in partner	this is the correct age of both male and female involved in the case	Legitimate; interviewer's override stands
63	Grandchild and respondent less than 30 years apart	45	16-year-old grandchild	This is step granddaughter	Legitimate; interviewer's override stands
64	Gap of 25 or more years between sister and respondent	66	41-year-old sibling	AGE IS CORRECT R MOM HAD THREE CHILDREN LATER IN LIFE	Legitimate; interviewer's override stands
65	Respondent's daughter-in-law is older than respondent	25	38-year-old child-in-law	daugher in law is simply older then wife of father	Legitimate; interviewer's override stands
66	Respondent is 16 or younger & married or has a live-in partner	16	17-year-old live-in partner	17 year old is R boyfriend and pregnant by him	Legitimate; interviewer's override stands
67	Respondent is 16 or younger & married or has a live-in partner	16	17-year-old live-in partner	have a baby together,now living together w/ mother until own home is ready	Legitimate; interviewer's override stands
68	Respondent's husband is 16 or younger	18	16-year-old live-in partner	Unmarried live in partner of 18 yr. old	Legitimate; interviewer's override stands
69	Respondent is 16 or younger & married or has a live-in partner	16	18-year-old live-in partner	Mother of R's baby girl is his livein girlfriend	Legitimate; interviewer's override stands

Table J.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
70	Grandparent and respondent less than 30 years apart	21	50-year-old grandparent	mother is 36 and grandma is 50	Legitimate; interviewer's override stands
71	Grandparent and respondent less than 30 years apart	21	50-year-old grandparent	grandpa is 50 mother 36 he is 21	Legitimate; interviewer's override stands
72	Respondent's son is older than respondent	52	55-year-old child	the stepson is older than the wife	Legitimate; interviewer's override stands
73	Grandparent and respondent less than 30 years apart	18	45-year-old granparent	Respondent still says grandfather is 46, she is 18	Legitimate; interviewer's override stands
74	Grandchild and respondent less than 30 years apart	60	31-year-old grandchild	STEP GRANDSON	Legitimate; interviewer's override stands
75	Respondent's wife is 16 or younger	17	16-year-old live-in partner	R living with pregnant girlfriend	Legitimate; interviewer's override stands
76	Grandparent and respondent less than 30 years apart	21	50-year-old grandparent	His father is 30ish, grandfather was young when his father born	Legitimate; interviewer's override stands
77	Grandparent and respondent less than 30 years apart	31	58-year-old grandparent	only 27 years different mother was 14 years old when r was born	Legitimate; interviewer's override stands
78	Gap of 25 or more years between brother and respondent	35	60-year-old sibling	R said it is correct	Legitimate; interviewer's override stands
79	Respondent is 16 or younger & married or has a live-in partner	16	18-year-old live-in partner	is her financee	Legitimate; interviewer's override stands

[1] These entries came directly from the 2005 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and not corrected.

[2] This consistency check was triggered because the roster member's age was originally reported as 5 years old instead of 5 months old, so the roster member's age was changed accordingly.