2000 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

IMPUTATION REPORT

Contract No. 283-98-9008
RTI Project No. 7190

Authors:

Eric A. Grau
Katherine R. Bowman
Elizabeth Copello
Peter Frechtel
Amy Licata
Dawn M. Odom

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI
Research Triangle Park, North Carolina 27709

July 30, 2002

2000 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

IMPUTATION REPORT

Contract No. 283-98-9008
RTI Project No. 7190

Authors:

Project Director: Thomas G. Virag

Eric A. Grau
Katherine R. Bowman
Elizabeth Copello
Peter Frechtel
Amy Licata
Dawn M. Odom

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI
Research Triangle Park, North Carolina 27709

July 30, 2002

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 Butler, Joe Gfroerer, and Art Hughes. Special thanks are also due to several RTI 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 7.3.4.3 and 7.3.4.4. Kortnee Barnett provided many helpful comments on Chapter 8. Finally, Richard Straw copyedited and 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)

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Appendix

List of Tables

List of Exhibits

Exhibit **Page**

List of Exhibits (continued)

List of Exhibits (continued)

List of Exhibits (continued)

Exhibit **Page**

List of Exhibits (continued)

List of Exhibits (continued)

Exhibit **Page**

List of Exhibits (continued)

Exhibit **Page**

List of Exhibits (continued)

Exhibit **Page**

1. Introduction

Starting in 1999 and continuing through 2000, the National Household Survey on Drug Abuse (NHSDA) was implemented using a new 50-State design. Other major changes in 1999 to the study protocol included the introduction of computer-assisted interviewing (CAI) methods for both screening households and interviewing selected respondents. An interview using paper-and-pencil interviewing (PAPI) methods was also included in 1999 for consistency with previous years. However, in 2000, only a CAI sample was selected.

The 50-State design was developed 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. This resulted in a major increase in sample size at the national level (from about 20,000 to 70,000 per year).

In 1999, 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. 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 interview 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).

A major objective of introducing CAI technology was to improve the quality of the data by providing valid substance use reports and by avoiding the inconsistencies that arise naturally in the PAPI approach. Under PAPI, sensitive sections of the interview were completed on separate answer sheets by the respondent. Instead of being instructed to follow skip instructions around nonapplicable questions, the respondent was asked to respond to each question, but also was allowed the option of indicating that a question did not apply. The CAI interview was programmed to automatically route the respondent to appropriate sections based on responses to gate questions, where "gate" refers to the first in a series of questions about a drug and indicates whether the respondent had ever used that drug. In addition, a number of consistency checks were programmed into the interviewing process to detect inconsistent answers and solicit the respondent's answers to additional questions intended to resolve the inconsistencies. Two of the expected benefits of the CAI approach included (a) more complete responses (fewer missing items) and (b) more internal consistency among responses to different questions.

The focus of this report is on procedures implemented for the 2000 NHSDA. The eligibility and completeness criteria are discussed in **Chapter 2**, followed by a summary of the implemented imputation procedures (**Chapter 3**). **Chapters 4** and **5** detail the imputation procedures applied to the core and noncore demographic variables, respectively. The drug imputation procedures are discussed in **Chapter 6**. Most of the editing procedures that were applied to the demographic and drug variables discussed in **Chapters 4**, **5**, and **6** are summarized by Kroutil (2002a, 2002b, 2002c). The editing procedures for the income and household composition variables, however, are discussed in this document. **Chapter 7** summarizes the editing and imputation procedures applied to the health insurance and income variables, and **Chapter 8** details 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.

This document also contains nine appendices, including three 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 new methodology developed specifically for the NHSDA, predictive mean neighborhoods (PMN), 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. These responses, called "alpha-specify" responses, were coded so that the data could be summarized in a meaningful way. A discussion of how this was done for race-Hispanicity and employment status is described in **Appendices D** and **E**. (Coding of alpha-specify responses for other variables is summarized by Kroutil, 2002a, 2002b, 2002c.) A summary of the models used in the PMN methodology for various variables is given in **Appendix F**.

A summary of the number of respondents who met various constraints that could be loosened in the imputation process is provided in **Appendix G**. **Appendix H** gives details of the vector of predicted means used in the multivariate PMN procedure for drugs and health insurance for various patterns of missing values, in addition to the logical constraints required. The quality control procedures implemented for the drug use imputations are summarized in **Appendix I**. For the 2000 NHSDA questionnaire specifications for programming, refer to RTI (2002).

2. Eligibility and Completeness Rules

2.1 Eligibility Criteria

The population of eligible respondents for the 2000 NHSDA was all civilian, noninstitutionalized residents of the United States (including the District of Columbia) aged 12 or older. As in other recent NHSDAs, this population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Persons excluded from the 2000 survey included those with no fixed household address (e.g., homeless transients *not* in shelters), residents of institutional group quarters, such as jails and hospitals, and active military personnel.

During screening, respondents were asked to identify all eligible household members so that only eligible individuals were listed and therefore potentially selected. Due to screening errors, some ineligible individuals were selected, however, and later determined to be ineligible at the time of interview. For a summary of the number of eligible persons rostered and completed interviews obtained in the 2000 NHSDA, see **Table 1**.

Table 1. Household and Person Eligibility and Response Rates, 2000 NHSDA

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases
CAI	215,860	182,576	169,769	349,673	91,961	71,993	71,764

2.2 Completed Case Rule

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 PAPI questionnaire in 1999 and prior NHSDAs, no logical inference could be made from information within a section if the gate question was not answered because the CAI instrument routed respondents out of a section if the gate question was not answered. For a summary of the number of completed cases in the 2000 NHSDA, see **Table 1**.

3. Overview of Item Imputation Procedures

3.1 Introduction

As with most large-scale sample surveys, the 2000 NHSDA 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 would be expected to reduce inconsistencies, and perform some consistency checks, inconsistent and invalid responses still occurred. These response errors are an obvious source of bias that must be considered in the analysis of NHSDA 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 are often replaced with statistically imputed data.

Unlike past years, where PAPI alone or a combination of PAPI and CAI were used, only CAI was used for the 2000 NHSDA. To maintain consistency with the 1999 NHSDA, some of the procedures implemented in the 2000 sample were identical to those used with the 1999 CAI. On the other hand, differences between the CAI instruments of 1999 and 2000 required modifications in the imputation procedures. In addition, a new procedure developed in 1999 specifically for the NHSDA, predictive mean neighborhoods (PMN), was applied to more variables in 2000. For other variables where PMN was already in place, the imputation was expanded from a univariate to a multivariate imputation. Other improvements in procedures were also implemented in 2000. **Exhibit 1** provides a brief summary of the types of imputation procedures used for each of the variables imputed in the samples from 1999 and 2000. This chapter provides brief descriptions of the two imputation procedures most used in the 2000 NHSDA, unweighted sequential hot deck and PMN, followed by a summary of the changes in imputation procedures from 1999 to 2000.

Exhibit 1. Summary of Item Imputation Procedure Used, by Variable and NHSDA Survey Year

Variable	1999 CAI	2000 CAI
Interview Date	Random[1]	Random
Age	None[2]	None
Birth Date	None	Random
Gender	None	None
Race	HD[3]	MPMN[4]
Hispanic-Origin Indicator	HD	UPMN[5]
Marital Status	HD	MPMN
Hispanic-Origin Group	HD	MPMN
Education	HD	HD
Employment Status (5 levels)	HD	HD
Private and Total Health Insurance	MPMN	MPMN
Drug Lifetime Usage (enters into recency)	UPMN	MPMN
Drug Recency of Use	MPMN	MPMN
ALC, MRJ, COC Frequency-of-Use (12 months)	MPMN	MPMN
Other Drug Frequency-of-Use (12 months)	MPMN	MPMN
Drug Frequency-of-Use (30 days)	MPMN	MPMN
Binge Drinking Frequency (30 days)	MPMN	MPMN
Age at First Use	UPMN	UPMN
Age at First Daily Cigarette Use	UPMN	UPMN
Personal and Family Income Binary Variables	MPMN	MPMN
Personal and Family Income Finer Categories	UPMN	UPMN
Household Size (Roster-Derived Variable)	UPMN	UPMN
Other Household Composition (Roster-Derived) Variables	UPMN	UPMN
Pair relationship variables and multiplicity/household counts	PMN[6]	PMN

1 "Random" refers to a random assignment within quarter for interview date, and a random assignment using age and interview date for birthdate.
2 "None" means that no missing values were encountered after editing, so that no imputation was necessary. For age, missing values are precluded by design (see **Chapter 4**).
3 "HD" refers to the unweighted sequential hot-deck method of item imputation described in this report (**Appendix A**).
4 "MPMN" refers to the "multivariate predictive mean neighborhood" model-based procedure described in this report (**Appendix C**).
5 "UPMN" refers to the "univariate predictive mean neighborhood" model-based procedure described in this report (**Appendix C**).
6 "PMN" refers to the "predictive mean neighborhood" model-based procedure that could be univariate or multivariate, depending upon the response variable of the model.

3.2 Overview of Unweighted Sequential Hot-Deck Imputation Procedure for the NHSDA Sample

The unweighted sequential hot-deck method of statistical item imputation was used for the 2000 NHSDA for educational level and employment status. In 1999, variables associated with race, Hispanic origin, and marital status were imputed using the unweighted sequential hot-deck method. Yet, in 2000, a PMN model-based imputation procedure was implemented for these variables. Educational level and employment status were not upgraded to the PMN method because of time constraints, low nonresponse, and both variables having large numbers of answer categories.

Simply, in the unweighted sequential hot-deck procedure, observations are sorted and classed by variables related to the variable of interest. This procedure involves replacing a missing value with a valid value taken from another respondent who is "similar" and has complete data. Those individuals who respond to the variable of interest are known as item respondents. Responding and nonresponding cases were sorted together by a variable or collection of variables closely related to the variable of interest (Y) that are known for both item respondents and item nonrespondents. When the item respondent and item nonrespondent are required to have the same values of a sorting variable, the variable is called a "classing variable." For sequential hot-deck imputation, a missing value of Y is replaced by the nearest responding value preceding it in the sequence. A complete discussion of the hot-deck method of imputation is presented in **Appendix A**.

3.3 Overview of PMN Imputation Procedure for the NHSDA Sample

PMN was developed specifically for the 1999 NHSDA. A combination of model-based imputation and a random nearest neighbor hot deck, PMN was implemented for the drug use, health insurance, and income variables, as well as the core demographic variables associated with race, Hispanic origin, and marital status. Missing values in all the core demographic variables were imputed using unweighted sequential hot deck in 1999.

When large nonresponse occurs, limited donor sets can be used for imputation. In 2000, to adjust for this sparseness of data, predictive mean modeling was used for the imputation of many of the variables (**Exhibit 1**). Models incorporated sampling design weights,[1] with a response propensity adjustment computed to make the item respondent weights representative of

---

[1] In the 2000 NHSDA, the final analysis weights were used if they were available. However, because the modeling of the final nonresponse adjustment 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.

the entire sample. The item response propensity model is a special case of the generalized exponential model (GEM), 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 (predictive 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 give a summary of the effects of covariates on the outcome variable. Unlike the unweighted sequential hot-deck imputation method, where values for the covariates are matched through a sorting procedure, the model-based approach uses the predictive mean to convert the covariates' effects into a single number. The predictive means, along with other constraints, are used to define the neighborhoods, from which donors were randomly selected for the final assignment of imputed values. This assignment was either done one value at a time (UPMN) or using several response variables at once (MPMN). More details regarding these UPMN and MPMN imputation procedures are given in **Appendix C**.

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 cannot be loosened. The latter, called "likeness constraints," can be loosened if insufficient donors are 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.

Because drug use, as well as variables related to income, insurance, and household composition are highly correlated with age and to facilitate easier implementation of the procedures, the model building and final assignments of imputed values for all drug, income, insurance, and household composition (roster-derived) variables were each done separately within distinct age groups. The drug variables were imputed within each of three age groups: 12 to 17 year olds, 18 to 25 year olds, and persons 26 years of age or older. The income, insurance, and household composition (roster-derived) variables were done within the following age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and persons 65 years of age or older. The age group restriction on the neighborhoods could be considered a likeness constraint. However, this restriction was never loosened because the models were also built separately for the age groups.

Although statistical imputation of the drug use variables could not proceed 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 of the PMN procedure in the 2000 CAI sample. 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. 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 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.

3.4 Changes in Procedures from 1999 to 2000

Overall, the changes implemented between 1999 and 2000 were minor, both in number and in kind. Some of these changes were the result of modifications to the instrument. Others, however, were enhancements that were implemented as a result of a review of the procedures used in 1999 and involved both editing and imputation.

3.4.1 Differences Between Instruments in 1999 and 2000 Affecting Variables Requiring Imputation

In 1999, a single set of choices was given for the race questions, which included each of the Asian subcategories (Chinese, Japanese, etc.). In the 2000 CAI instrument, a single category for Asians was given in the main race question, then Asians were given the opportunity to list the appropriate, more descriptive subcategory. As a result, if the interviewer needed to manually enter a category not covered by those in the race question(s), separate opportunities to do so were provided for Asian and non-Asian respondents' races. Although this did not have a direct impact on the imputation of race or Hispanicity, it did affect the editing of variables related to race and Hispanicity, which is discussed in **Chapter 4**.

For health insurance, two new questions were asked about overall health insurance. This provided an opportunity to create a new variable for 2000, INSUR2, which incorporated the information obtained from the extra questions. Details are given in **Chapter 7**.

3.4.2 Improvements in Imputation Procedures

All demographic variables were imputed using the unweighted sequential hot-deck procedure (Little & Rubin, 1987) in 1999, which is summarized in **Section 3.3**, and discussed in detail in **Appendix A**. This was due to the low amount of missingness in all of these variables (except for race) and the limited amount of time available to implement the procedure. With the exception of education and employment, missing values in all demographic variables in 2000 were imputed using the PMN method, which is summarized in **Section 3.3** and discussed in detail in **Appendix C**. Details of these changes are summarized in **Chapter 4**.

Improvements to the drug use imputations were also implemented in 2000. For instance, in 1999, provisional values for drug lifetime usage were calculated using an unweighted sequential hot-deck imputation method; in 2000, these provisional values were obtained using a univariate predictive mean neighborhood (UPMN) imputation method. Whereas the final imputed values for lifetime usage were obtained using a series of UPMNs in 1999, the final imputed values in 2000 were obtained using a multivariate predictive mean neighborhood (MPMN) imputation method. These changes are summarized in **Chapter 6**.

3.4.3 Other Improvements in Procedures from 1999 to 2000

In 1999, the variable NEWRACE1, which includes levels for Hispanicity and detailed race categories, was created late in the processing schedule, after the Hispanic indicator IRHOIND and the four-level race variable IRRACE had already been created. To maintain consistency with these variables, the detailed race categories for NEWRACE1 were obtained using information from the IRRACE donors. In some cases, a supplemental imputation was performed to fill in gaps in the information provided by those donors. (Details are given in the 1999 imputation report.) In 2000, the creation of a new race variable, IRNWRACE, obviated this supplemental imputation. IRNWRACE, which includes detailed race categories without a level for Hispanicity, was created simultaneously with IRRACE. The variable NEWRACE1 was then created as a recode using the information from IRNWRACE and IRHOIND.

A number of improvements were implemented on the household roster edits. Many of these improvements were related to the effort to more easily establish the relationship between selected pair members. To increase the usability of the household roster, individual roster-level variables were created for the first time in the 2000 data. Greater details regarding the editing of the household composition variables are presented in **Chapter 8**.

Finally, more thorough editing was performed on several variables, including INTDATE, AGE, and BRTHDATE. More details are available in **Chapter 4**. Even though quality control was executed in 1999, one notable improvement in 2000 was the increased amount of checks that were implemented to ensure that the imputations procedures were yielding the desired results and were being performed appropriately. Greater details about the specific quality control standards that were used in 2000 can be found in **Appendix I**.

4. Core Demographics

4.1 Introduction

As in previous NHSDAs, several demographic characteristics were needed for all respondents in the 2000 NHSDA. Core demographic data were collected on both the screener and the questionnaire and 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).[2] 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 2000 NHSDA discussed in this document are age, birth date, gender, race, Hispanic origin, marital status, and educational level (highest grade completed). The only noncore demographic variable imputed was employment status. Although the interview date is not classified as a core demographic variable, its editing procedures are also included in this chapter.

Prior to imputation, logical editing was performed on all of these variables. Through the editing process, some missing values are supplied, thus reducing the amount of statistical imputation required.[3] Logical editing of variables was done using only the "other-specify" questionnaire responses, and no noncore information was used to edit core variables.

After editing, the variables were handled using one of three procedures. For age, birth date, and gender, no statistical imputation was required because no values were missing after editing. Interview date had only one missing value, which was imputed using a random assignment within the quarter. The missing values in the marital status, race, and Hispanicity variables were imputed using a predictive mean neighborhood (PMN) method. This procedure is described in greater detail in **Appendix C**. For educational level, missing values were imputed using an unweighted sequential hot-deck procedure. Missing values for the noncore employment status variables, which are discussed in the next chapter, were also imputed using an unweighted sequential hot deck. This procedure is described in greater detail in **Appendix A**.

---

[2] See the *2000 NHSDA: Sample Design Report* for a description of the imputation procedures used for screener demographics for the set of all eligible rostered individuals (Bowman, Penne, Chromy, & Odom, 2002).

[3] Logical editing undertaken to create base variables for imputation is described in this report; for more details on other editing performed on NHSDA data prior to imputation, see Kroutil (2002a, 2002b, 2002c).

This chapter describes the editing and imputation procedures used to create final demographic variables for all respondents. A summary of item nonresponse is included for each variable described here.

4.2 Variables Commonly Used as Covariates or as Sorting/Classing Variables

In the PMN procedure, statistical modeling is done both to adjust weights for item nonresponse and to calculate predictive means. In the unweighted sequential hot-deck procedure, the observations are sorted and classed by variables related to the variable of interest. The following variables were often used as covariates in both types of models for the PMN procedures, and/or as sorting and classing variables for educational level imputation using the hot-deck method.

4.2.1 Household Type

Household type is a three-level race/ethnicity variable based on screener data. It is created by recoding the race/ethnicity of the screening head of household to one of three levels: Hispanic, non-Hispanic black, or non-Hispanic non-black.

4.2.2 Region

Region is a four-level geographic variable recoded from the respondent's State of residence. The four levels are Northeast, Midwest, South, and West.

4.2.3 Segment ID

As described in the *2000 NHSDA: Sample Design Report* (Bowman et al., 2002), States were partitioned into field interviewer regions ("FI regions"), which were further partitioned into clusters of adjacent blocks called "segments." The variable SEGID (segment ID number) is a two-letter State abbreviation followed by a two-digit FI region and two-digit segment identifier, uniquely identifying each segment. SEGID was used as a sorting variable in the imputation of educational status. For more information regarding segments, see Bowman et al. (2002).

4.2.4 Population Density

The population density variable PDEN2 was used to categorize segments according to 1990 Census data adjusted to more recent data from Claritas, Inc. PDEN2 has 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.

4.2.5 Percent Hispanic Population

The Hispanic population variable HISPCONC was also used to categorize segments according to adjusted 1990 census data. It has three levels: less than 20 percent, 20 to 70 percent, and more than 70 percent.

4.2.6 Percent Non-Hispanic Black Population

The non-Hispanic black population variable NHBPCONC was also used to categorize segments according to adjusted 1990 Census data. It also has three levels: less than 10 percent, 10 to 50 percent, and 50 percent or more.

4.2.7 Percent of Owner-Occupied Households

The owner-occupied household variable OWNOCONC was also used to categorize segments according to adjusted 1990 Census data. It is used as a surrogate for income because wealthy segments tend to have many homeowners, and poor segments tend to have many renters. It has three levels: less than 10 percent, 10 to 50 percent, and 50 percent or more.

4.3 Preliminary Edits: Interview Date, Age, and Birth Date

In the sample, the date of the interview, age, and birth date were required for all completed cases. Some editing of these date values was required to resolve inconsistencies and fill in missing data. The editing of all three variables, especially the interview date and age, was much more thorough in 2000 than in 1999. This was due to the fact that subsequent analyses depend crucially upon having an accurate value for age and interview date. This meant that more effort was required to determine the best estimates of these variables. For example, to determine a better estimate of interview date, RTI contacted the field interviewer (FI) if the given interview date was outside the quarter. In 1999, for age, it was considered acceptable to have a value within 1 year of a "better" estimate. In 2000, considerable effort was exerted to determine if another value of age was better.

4.3.1 Edited Interview Date (INTDATE)

Within each module of the questionnaire, the time that the module was filled out is automatically saved by the CAI instrument. The time for each module is called a "time stamp," and the date portion of the time stamp is called a "date stamp." In 1999, the date stamp information was not used in the determination of interview date. However, in 2000, 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) are indicated in the variable EIIDATE. For the labels that define the levels in EIIDATE, if the label indicates that the interview date was set to a particular date stamp, that date stamp is consistent with all subsequent date stamps unless otherwise indicated. If the interview is set to the end-of-interview date stamp, that date stamp is 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 is indicated).

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

- The date stamp was 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.

If no other source of information about the interview date was available, the interview date was randomly assigned within the correct quarter. For a summary of the editing of interview dates, see **Table 2**. This information is recorded in the editing indicator variable EIIDATE.

Table 2. Interview Date Editing Summary

Value of EIIDATE	Assignment of Interview Date	Frequency	Percent
1	Begin date stamp (all date stamps exist)	71623	99.80
1.01	Begin date stamp (all date stamps exist except last one)	19	0.03
1.02	Begin date stamp (all date stamps exist up through sedatives)	39	0.05
1.03	Begin date stamp (all date stamps exist up through stimulants)	2	0.00
1.04	Begin date stamp (all date stamps exist up through tranquilizers)	2	0.00
1.06	Begin date stamp (all date stamps exist up through inhalants)	5	0.01
2	Last existing date stamp (earlier than begin date stamp)	3	0.00
3	Tutorial date stamp (begin date stamp outside quarter)	5	0.01
4	End date stamp (begin and tutorial date stamps outside quarter)	1	0.00
5	Beginning date stamp with corrected year	11	0.02
6	Date later manually entered from RTI investigation	35	0.05
7	Tutorial date stamp (begin date stamp missing)	3	0.00
8	End date stamp (tutorial date stamp first after birthday)	3	0.00
8.01	End date stamp (cigarettes date stamp first after birthday)	1	0.00
8.03	End date stamp (snuff date stamp first after birthday)	4	0.01
8.05	End date stamp (pipes date stamp first after birthday)	1	0.00
8.08	End date stamp (cocaine date stamp first after birthday)	1	0.00
8.16	End date stamp (not others due to birthday)	5	0.01
9	Randomly assigned within quarter	1	0.00

4.3.2 Age

4.3.2.1 Final Edited Continuous Age (AGE)

After a respondent has entered his/her birth date in the first part of the questionnaire, he/she has multiple opportunities to change his/her age in response to consistency checks throughout the questionnaire. It is therefore possible for the age recorded by the respondent at the beginning of the questionnaire (CALCAGE) to be different from the age at the end of the questionnaire (NEWAGE). The final age variable, AGE, is determined using these two variables, in addition to the age calculated from the raw birth date (AGE1) and the final edited interview date (INTDATE), the age entered in the questionnaire roster (if it exists), and the pre-interview screener age. When determining the final edited continuous age, priority is given to CALCAGE, NEWAGE, and the age calculated from AGE1 and INTDATE. If the final age (AGE) does not agree with the originally entered birth date (AGE1), the birth date must also be edited. The final edited variable AGE was determined in the following manner:

> 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 are not equal, but NEWAGE is exactly equal to CALCAGE (adjusted by Blaise[4] to a changed interview date if the interview date was changed within the questionnaire, and the respondent's birthday does not fall between the dates corresponding to TBEG_TUT and INTDATE (age indicator = 1), else
>
> NEWAGE, if nonmissing, TBEG_TUT and INTDATE are not equal, the respondent's birthday falls between the dates corresponding to TBEG_TUT and INTDATE, the given value of CALCAGE agrees with what it should be based on INTDATE and the given birth date (i.e., EIIDATE not equal to 6), and NEWAGE and CALCAGE are exactly equal (age indicator = 1), else
>
> age calculated from INTDATE and the reported birth date, if the birth date is nonmissing, TBEG_TUT and INTDATE are not equal, the respondent's birthday falls between the dates corresponding to TBEG_TUT and INTDATE, the given value of CALCAGE does 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 is exactly equal to the screener age and the roster age (if it exists) (age indicator = 2), else
>
> NEWAGE, if NEWAGE differs from CALCAGE and NEWAGE = screener age and NEWAGE = roster age (if it exists), and the interview date at the beginning of the interview (TBEGINTR) is within the appropriate quarter (age indicator = 3), else
>
> CALCAGE, if CALCAGE differs from NEWAGE and CALCAGE = screener age and CALCAGE = roster age (if it exists), and the interview date at the beginning of the interview (TBEGINTR) is within the appropriate quarter (age indicator = 4), else
>
> age calculated from reported birth date and INTDATE, if EIIDATE = 5 and NEWAGE = CALCAGE (but neither are equal to the correct age) (age indicator = 5), else

---

[4] Blaise is the computer program that performs the automatic skips within the questionnaire laptop computer.

NEWAGE, if NEWAGE differs from CALCAGE, but NEWAGE = roster age, provided roster age exists (age indicator = 6), else

CALCAGE, if CALCAGE differs from NEWAGE, but CALCAGE = roster age, provided roster age exists (age indicator = 7).

For a summary of the editing to create AGE for the 2000 NHSDA, see **Table 3**. This information is recorded in the editing indicator variable EIAGE.

Table 3. Age Editing Summary

Value of EIAGE	Assignment of Age	Frequency	Percent
1	NEWAGE (consistent with CALCAGE and INTDATE - AGE1)	71735	99.96
2	Age from INTDATE and AGE1 (consistent with screener age)	9	0.01
3	NEWAGE (consistent with screener age)	4	0.01
4	CALCAGE (consistent with screener age)	5	0.01
5	Age calculated from INTDATE and AGE1; INTDATE was assigned the beginning date stamp with corrected year	7	0.01
6	NEWAGE (consistent with roster age)	1	0.00
7	CALCAGE (consistent with roster age)	3	0.00

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

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

4.3.3 Edited Birth Date (BRTHDATE)

Respondents were required to provide their date of birth and/or current age at the beginning of the interview in order to continue with the questionnaire. 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, the birth date was calculated as follows:

- The integer value from the final edited age was converted to a SAS date value[5] by first adding a fraction of a year (in the form of a randomly generated fraction) and then multiplying by 365.25:

Intermediate age = [*Final edited age* + *Uniform* (0,1) *number*] * 365.25.

- The final birth date was set equal to the difference between the edited interview date and the intermediate age variable (a SAS date value):

BRTHDATE = Edited interview date - *Intermediate age*.

See **Table 4** for a summary of the birth date editing. This information is recorded in the editing indicator variable EIBDATE.

Table 4. Birth Date Editing Summary

Value of EIBDATE	Assignment of Birth Date	Frequency	Percent
1	Reported birth date	71675	99.88
2	Reported birthday, year from AGE and INTDATE	8	0.01
3	Randomly assigned using AGE and INTDATE	81	0.11

4.4 Demographics Requiring Imputation

Missing values for the demographic variables of completed cases were imputed separately from those of all eligible (screener) rostered individuals. Moreover, no screener information was used to edit questionnaire demographics for the completed cases, except in some extraordinary circumstances, which are explained below. The descriptions that follow discuss the creation of edited and imputed demographic variables. However, the edited variables are entirely internal; only imputed variables were released to the analytic and public use files.

4.4.1 Gender

4.4.1.1 Edited Gender (EDSEX)

An edited gender variable (EDSEX) was created for all respondents. For the vast majority of cases, EDSEX was simply set equal to the gender reported by the respondent in response to question QD01. When no gender was reported in the questionnaire, EDSEX was set

[5] SAS date values are stored as the number of days since January 1, 1960.

equal to the gender reported during the household screening.[6] In the sample of the 2000 NHSDA, there were no additional missing values for EDSEX; therefore, no statistical imputation was required. For a summary of item nonresponse and editing for gender, see **Table 5**. This information is recorded in the imputation indicator variable IISEX.

Table 5. Gender Editing and Imputation Summary

Value of IISEX	Assignment of Gender	Frequency	Percent
1	From questionnaire	71761	100.00
2	From screener gender	3	0.00
3	Statistically imputed	0	0.00

4.4.1.2 Imputation-Revised Gender (IRSEX)

The final version of the gender variable was called IRSEX. In 2000, no statistical imputation was required to create this variable because gender was determined from questionnaire or screening responses for all respondents. For a summary of item nonresponse for gender, see **Table 5**.

4.4.2 Race

4.4.2.1 Edited Race (EDRACE) and Edited Race, Finer Categories (NEWRACE)

In the 2000 questionnaire, two core questions focused on the respondent's race. The first question (QD05) allowed the respondent to select multiple race categories, and the second (QD06) asked the respondent to choose from among those selected in QD05 (if more than one race was selected) the single race that best describes him or her. QD05 has an "other" category, and if this category was chosen in QD05, the respondent was asked to specify a race; this "alpha-specify response" appeared as a possible response in QD06. Unlike in the 1999 CAI questionnaire, where the specific "Asian" categories were possible responses in QD05, in 2000, the "Asian" response to QD05 routes the respondent to a separate question with more specific Asian ethnic groups. This separate question (QD05ASIA) also has an "other" category, which is treated the same way as the "other" category in QD05. Even though the specific Asian categories

[6] It was a policy not to use the screener to edit questionnaire responses because the variables collected on the screener would vary from 1 year to the next, and the person giving the screening information about the respondent may not in fact be the respondent. However, the number of missing values for gender was very small, and the quality of the imputed values for gender would probably be very low. Hence, an exception was made for gender to obtain values from the screener when questionnaire values were unavailable.

appeared in an additional question in 2000, the answers to QD05ASIA were treated exactly as if they came from QD05.[7]

There were 13 answer categories in the race questions. However, the final race variable IRRACE was a four-level nominal variable, which was collapsed to the same levels as those of past NHSDAs: American Indian or Alaska Native, Asian or Pacific Islander, black, and white.[8] The final finer race-categories variable IRNWRACE has the 13 answer categories, plus "Asian multiple category" and "more than one race."

EDRACE, the base variable for imputing race, was created as follows. If only one race was chosen in response to QD05, EDRACE =

the single race identified in QD05, if that single race was not "other," else

race recode from alpha-specify response(s)[9] when "other" or "other Asian" was the only race selected in QD05, if a valid recode was available,[10] else

missing.

If more than one race was chosen in response to QD05, EDRACE =

the race response in QD06, if it is not "other," "other Asian," or missing, else

---

[7] The only other change to the 2000 questionnaire from the 1999 CAI was a reordering of the answer categories. In the SAS code, the answers were rearranged to be in the same order as in 1999, and the rest of the code ran as in 1999. One issue that arose in 2000 but not for 1999 was an accidental mapping of Asian Indians to the American Indian/Alaska Native category, which was corrected in the code. This occurred because the same alpha-specify dictionary was used for the "other" and "other Asian" categories; if a respondent chose "other Asian" in QD05ASIA but mentioned "Indian" in their alpha-specify response OTHASIA, he or she was originally mapped to "American Indian/Alaska Native" by the dictionary.

[8] 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.

[9] QD04 (Hispanic-origin group question, see **Section 4.4.6**), QD05, and QD05ASIA allowed interviewers to enter a written response to the questions about the respondent's Hispanic group or race, respectively, when the listed responses were seen not to apply and the category "other" was selected. These written responses are called "alpha-specify" responses, which were coded using the lookup table given in **Appendix D**. In many cases, respondents keyed in a racial category in response to the Hispanic-origin group question (QD04) or a Hispanic origin group in response to the race question(s) (QD05 or QD06). Thus, in checking alpha-specify responses for the race and Hispanic-origin group variables, both QD04 and QD05 were checked for each category. For a detailed description of the assignment of race categories from alpha-specify responses, see **Appendix D**.

[10] In a number of cases, the race and/or Hispanic origin group specified by a respondent did not fit into the categories used by NHSDA, or the respondent did not specify a race when prompted, so no recode was available (see **Appendix D**).

race recode from alpha-specify response if QD06 = "other" or "other Asian" and a valid recode is available, else

race assigned from the multiple responses given to QD05, using the following priority: black/African American, Asian, American Indian/Alaskan Native, white.[11]

If no response was given to QD05, EDRACE =

race recode from alpha-specify response to QD04 (Hispanic origin group), if a valid recode is available, else

missing.

NEWRACE, the base variable for imputing the finer categories race variable, was created in the following manner:

NEWRACE = missing, if QD05 was missing, or if QD05 = "other" and/or "other Asian", and the alpha-specify response(s) was Asian, but not one of the specific Asian categories included in QD05ASIA, else

1 (White only), if either white was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "white," else

2 (Black/African American only), if either black/African American was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "black/African American," else

3 (Native American or Alaska Native), if either Native American was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Native American," else

4 (Native Hawaiian only), if either Native Hawaiian was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Native Hawaiian," else

---

[11] To select one racial group from multiple selected groups, a priority rule was established whereby if black/African American was among the groups selected, the single race for the respondent is black/African American; otherwise, if Asian was among the groups selected, the single race for the respondent is Asian, etc. Details are given in **Appendix D.**

5 (Other Pacific Islander only), if either Other Pacific Islander was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Other Pacific Islander," else

6 (Native Hawaiian and Other Pacific Islander), if both Native Hawaiian and Other Pacific Islander were selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Native Hawaiian and Other Pacific Islander," else

7 (Chinese only), if either Chinese was the only race selected in QD05, or "other" or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Chinese," else

8 (Filipino only), if either Filipino was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Filipino," else

9 (Japanese only), if either Japanese was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Japanese," else

10 (Asian Indian only), if either Asian Indian was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Asian Indian,"[12] else

11 (Korean only), if either Korean was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Korean," else

12 (Vietnamese only), if either Vietnamese was the only race selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "Vietnamese," else

13 (Other Asian only), if "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as "other Asian," else

14 (Asian multiple category), if either more than one race was selected in QD05 where all those selected are considered "Asian," or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was interpreted as a combination of several Asian categories, else

[12] When the respondent's alpha-specify response for "other Asian" was "Indian" or something similar, the respondent was determined to be Asian Indian, not American Indian. This occurred for seven respondents.

15 (Multiple Race), if two or more races were selected in QD05 and (a) at least one was non-Asian, and (b) at least one was something other than Native Hawaiian or Other Pacific Islander.

4.4.2.2 Imputation-Revised Race (IRRACE) and Imputation-Revised NEWRACE (IRNWRACE)

For the first time, imputation-revised race variables were created using a multivariate predictive mean neighborhood (MPMN) method for imputation of missing values. The PMN method as applied to the race variables is explained in detail in the next four sections: setup for model building, computation of predictive means, assignment of imputed values, and constraints on MPMNs.

4.4.2.2.1 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 year olds, 18 to 25 year olds, and respondents aged 26 or older. Because all interview respondents were asked the race questions, no subsetting of the data was necessary.

Before predictive mean modeling was implemented, analysis weights were adjusted for item nonresponse to the race questions. An interview respondent was considered an item nonrespondent for race if either EDRACE was missing, NEWRACE was missing, or both. The analysis weights of the item nonrespondents were redistributed among the item respondents using an item response propensity model, which is described in greater detail in **Appendix B**. Covariates used in the model were the same for each age group. These covariates were Census region, household type, final edited age, percent Hispanic population, percent non-Hispanic black population, and percent of owner-occupied households.

4.4.2.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each race category was modeled within each age group using polytomous logistic regression.[13] The predictors included in the models were the same as those used in the item response propensity model for race. Due to

[13] SAS-callable SUDAAN was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model can be found in the *SUDAAN User's Manual, Release 8.0* (RTI, 2001). Additional references are provided in this user's manual.

interest only in the estimation of the predictive mean, and not in the parameter estimates (by themselves) or their standard errors, no model selection was attempted.

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 falling into each of the first three race categories. The probability of falling into the fourth race category was not included because it is completely defined by the first three elements in the predictive mean vector: it is one minus their sum.

4.4.2.2.3 Assignment of Imputed Values

The PMN method requires the selection of an item respondent to the race questions who is similar to each item nonrespondent, who will "donate" his or her value for EDRACE to the item nonrespondent. Most often, the selected item respondent, called the "donor," is randomly chosen from a "neighborhood" of potential donors. The item respondents in this neighborhood are the ones deemed to be most similar to the given item nonrespondent, who is called the "recipient." Item respondents who are deemed dissimilar to the recipient are discarded from the neighborhood by means of constraints. The predictive means calculated in the previous step are usually considered in these constraints. Because multiple variables are considered in the distance measure, "similarity" is defined in terms of the smallest Mahalanobis distance.[14] 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 assignments is executed for all imputation-revised variables in the NHSDA. If the recipient had missing values for both EDRACE and NEWRACE, the donor gave values for both variables to the recipient. This ensures consistency between IRRACE and IRNWRACE.

4.4.2.2.4 Constraints on MPMNs

There are two types of constraints: logical constraints and likeness constraints. Logical constraints are not loosened during the search for a donor. Likeness constraints can be loosened or removed if a donor cannot be found with the given constraints in effect. The logical constraints on the donors for EDRACE and NEWRACE are listed below:

---

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

- If the recipient was of Hispanic origin according to QD03, the donor must have also answered QD03 affirmatively.

- If the recipient answered QD04 (Hispanic group question) with any of the four specified responses (Mexican, Puerto Rican, Central or South American, or Cuban), the donor must have given the same answer.

- If the recipient was known to be Asian from QD05, the donor must also have been Asian.

- If the recipient was known to be of multiple race, but the specific races were unknown, the donor must not have been white.

In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint states that the donor must have lived in the same segment as the recipient. The second likeness constraint states that each of the donor's three predictive means, as described in **Section 4.4.2.2.2**, must have been within 5 percent of each of the recipient's three predictive 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. A donor was found for every item nonrespondent using this method; therefore, no further loosening of constraints was necessary. 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**.

4.4.2.3 Imputation and Editing Summary for Race

Table 6 summarizes the response categories for race. The imputation indicator variable II2RACE contains this information. The imputation indicator IIRACE is only included on the master dataset because of the results of the 1999 imputation, which was performed using an unweighted sequential hot-deck method, not the PMN method used in 2000. The II2RACE variable is both more detailed and more tailored to the PMN method than the IIRACE variable.

Table 7 summarizes the response categories for NEWRACE, the base variable for the imputation-revised finer categories race variable IRNWRACE. The imputation indicator variable IINWRACE contains this information.

Table 6. IRRACE Editing and Imputation Summary

Value of II2RACE	Assignment of IRRACE	Frequency	Percent
1	From single QD05 response	67638	94.25
2	From QD06 response	1136	1.58
3	Logically assigned from alpha-specify response	693	0.97
4	Assigned with Census data from country of origin	136	0.19
5	Single race determined from multiple responses	73	0.10
6	Statistically imputed (unrestricted)	96	0.13
7	Statistically imputed (restricted to Hispanic groups)	1992	2.78

Table 7. IRNWRACE Editing and Imputation Summary

Value of IINWRACE	Assignment of IRNWRACE	Frequency	Percent
1	From QD05 response(s)	68829	95.91
2	Logically assigned from alpha-specify response(s)	8	0.01
3	Assigned with Census data from country of origin	811	1.13
4	Statistical imputation of "Asian" into finer categories	28	0.04
5	Statistically imputed (unrestricted)	96	0.13
6	Statistically imputed (restricted using Hispanicity)	1992	2.78

4.4.3 Hispanic Origin (Dichotomous Indicator)

4.4.3.1 Edited Hispanic-Origin Indicator (EDQD04 and EDHOIND)

Prior to creating an edited Hispanic-origin indicator, an edited version of QD04 (EDQD04) was created. If respondents indicated that they were Hispanic in response to QD03, QD04 asked them to indicate which Hispanic origin group best describes them. If QD04's "other" category was chosen, the respondent was asked to specify a Hispanic-origin group. Respondents had the option of selecting more than one Hispanic group in QD04, but the final imputed Hispanic-origin group variable was limited to one category.

EDQD04 was created as follows. If only one Hispanic-origin group was selected in QD04, EDQD04 =

QD04, if it is not "other," else

Hispanic-origin group recode from alpha-specify response(s),[10] if "other" was selected and a valid recode is available,[11] else

missing.

If more than one Hispanic group was selected in QD04, EDQD04 =

Hispanic-origin group assigned from among the categories selected in QD04, according to the following priorities: Mexican, Cuban, Puerto Rican, Central American or South American.

If no groups were selected in QD04, EDQD04 =

Hispanic-origin group recode from alpha-specify response to QD05, if a valid recode is available, else

missing.

The base variable for creating an imputation-revised Hispanic-origin indicator is EDHOIND, which was created using responses to QD03 and the edited Hispanic-origin group variable (EDQDO4) as follows:

EDHOIND = 1 (Hispanic), if QD03 = 1 OR if alpha-specify response to QD05 indicates that the respondent is Hispanic OR if EDQD04 has a value indicating that the respondent is Hispanic, else

2 (not Hispanic), if QD03 = 2 OR if alpha-specify response to QD05 indicates that the respondent is not Hispanic OR if EDQD04 = 10, indicating that the respondent is not Hispanic, else

missing.

---

[10] Both QD04 (Hispanic-origin group question) and QD05/QD06 allow respondents to specify a race or Hispanic-origin group, respectively, other than those listed in the questions, when they select the category "other." In many cases, respondents keyed in a racial category in response to the Hispanic-origin group question (QD04) or a Hispanic-origin group in response to the race question(s) (QD05 or QD06). Thus, in checking alpha-specify responses for the race and Hispanic-origin group variables, both QD04 and QD05 were checked for each. For a detailed description of the assignment of race categories from alpha-specify responses, see **Appendix D.**

[11] In a number of cases, the race and/or Hispanic-origin group specified by a respondent did not fit into the categories used by NHSDA, or the respondent did not specify a race when prompted, so no recode was available. See **Appendix D.**

4.4.3.2 Imputation-Revised Hispanic-Origin Indicator (IRHOIND)

As with the imputation-revised race variables, a PMN method was used for the Hispanic-origin 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-origin indicator is explained in detail in the next four sections: setup for model building, computation of the predictive mean, assignment of imputed values, and constraints on UPMNs.

4.4.3.2.1 Setup for Model Building

As with imputations for other race variables, the imputations for the Hispanic-origin indicator were conducted separately within the three age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents aged 26 or older. Because all interview respondents were asked the question about Hispanic origin, no subsetting of the data was necessary.

As for the race variables, analysis weights were adjusted for item nonresponse to the Hispanic origin question, QD03. The covariates in the item response propensity model (see **Appendix B** for the more general GEM) were Census region, imputation-revised race, age, age squared, percent Hispanic population, percent non-Hispanic black population, and percent of owner-occupied households.

4.4.3.2.2 Computation of the Predictive Means

Using the adjusted weights, the probability of an affirmative response to the Hispanic 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, age, age squared, age cubed, percent Hispanic population, percent non-Hispanic black population, and percent of owner-occupied households.

4.4.3.2.3 Assignment of Imputed Values

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

4.4.3.2.4 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 states that the donor must have lived in the same segment as the recipient. The second likeness constraint states that the donor's predictive mean, as described in **Section 4.4.3.2.2**, must have been within 5 percent of the recipient's predictive 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 meeting each set of likeness constraints on sets of eligible donors.

4.4.3.3 Imputation and Editing Summary for Hispanic Origin

Less imputation was required for the Hispanic-origin indicator than for the race variables. **Table 8** summarizes item nonresponse for the Hispanic-origin indicator. This information is recorded in the variable IIHOIND.

Table 8. Hispanic-Origin Indicator Editing and Imputation Summary

Value of IIHOIND	Assignment of IRHOIND	Frequency	Percent
1	From questionnaire	71610	99.79
2	From alpha-specify responses	106	0.15
3	Statistically imputed	48	0.07

4.4.4 Race and Hispanicity Recodes

The imputation-revised race (IRRACE) and imputation-revised Hispanic-origin indicator (IRHOIND) variables were used to create two additional race/ethnicity variables: HISPRACE with three levels (Hispanic, non-Hispanic black, and non-Hispanic nonblack) and RACE with four levels (non-Hispanic white, non-Hispanic black, Hispanic, and non-Hispanic other).

Furthermore, two additional race/ethnicity variables were created from the imputation-revised finer categories race (IRNWRACE) and IRHOIND. These variables, NEWRACE1 and NEWRACE2, provide more detail by including all of the race categories included in the question and by indicating whether a respondent selected more than one race category. These variables indicated whether a respondent was Hispanic, based on IRHOIND, and the detailed race information was provided only for non-Hispanic respondents.

First, Hispanic respondents were assigned the NEWRACE1 and NEWRACE2 values of "Hispanic" using IRHOIND. Then, IRNWRACE was used to assign values to NEWRACE1 and NEWRACE2 for non-Hispanic respondents.

NEWRACE1 = IRNWRACE, if IRHOIND = 1, else

16 (Hispanic), if IRHOIND = 1.

NEWRACE2 was created by collapsing some of the levels of NEWRACE1 as follows:

NEWRACE2 = 1 (non-Hispanic white only), if NEWRACE1 = 1, else

2 (non-Hispanic black/African American only), if NEWRACE1 = 2, else

3 (non-Hispanic Native American or Alaska Native only), if NEWRACE1 = 3, else

4 (non-Hispanic Native Hawaiian or Other Pacific Islander only), if NEWRACE1 = 4, 5, or 6, else

5 (non-Hispanic Asian only), if NEWRACE1 = 7, 8, 9, 10, 11, 12, 13, or 14, else

6 (non-Hispanic Multiple Races), if NEWRACE1 = 15, else

7 (Hispanic), if NEWRACE1 = 16.

4.4.5 Marital Status

4.4.5.1 Edited Marital Status (EDMARIT)

The base variable for creating an imputation-revised version of marital status was called EDMARIT and was created in the following manner:

EDMARIT = QD07, if nonmissing and the respondent is 15 years old or older, else

99 (legitimate skip) if the respondent is younger than 15, else

missing.

4.4.5.2 Imputation-Revised Marital Status (IRMARIT)

The MPMN method used for marital status was similar to the method for IRRACE, in that the variable of interest is a four-level nominal variable. The four substantive levels of the imputation-revised marital status variable, IRMARIT, are the same as the four answer categories for QD07: married, widowed, divorced or separated, and never married. Respondents younger than 15 were automatically assigned an IRMARIT value of 99, a "legitimate skip" code. As with race, the predictive mean vector had three elements (i.e., the predicted probability of the interview respondent falling into each of the first three answer categories). The main differences between marital status imputation and race imputation are the relative simplicity of the editing process (Kroutil, 2002a, 2002b, 2002c) and the smaller domain of the variable (interview respondents younger than 15 were eliminated from the imputation dataset and logically assigned 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 predictive means, assignment of imputed values, and constraints on MPMNs.

4.4.5.2.1 Setup for Model Building

Imputations were conducted separately within the same three age groups as for the other demographic variables. All respondents with AGE less than 15 were assigned IRMARIT=99, regardless of their value for EDMARIT. Only interview respondents with AGE of 15 or greater were considered as donors.

An interview respondent was considered an item nonrespondent for marital status if his/her value for EDMARIT was missing. The weights of the item nonrespondents 15 or older were reallocated to the item respondents 15 or older. The covariates in the item response propensity model were Census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, population density, age, percent Hispanic population, percent non-Hispanic black population, percent of owner-occupied households, and the first-order interaction of age and gender.

4.4.5.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each marital status category was modeled within each age group using polytomous logistic regression.[12] The predictors included in the models were Census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, population density, age, age squared, age cubed,

---

[12] See earlier footnote in **Section 4.4.2.2.2** where a reference for polytomous regression is given.

percent Hispanic population, percent non-Hispanic black population, percent of owner-occupied households, and the first-order interaction of age and gender.

4.4.5.2.3 Assignment of Imputed Values

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

4.4.5.2.4 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, one likeness constraint was used. This constraint required each of the donor's three predictive means, as described in **Section 4.4.5.2.2**, to be within 5 percent of each of the recipient's three predictive means. If no item respondents met the above conditions for a particular item nonrespondent, the constraint on the predictive means was removed. See **Appendix G** for the numbers of respondents meeting each set of likeness constraints on sets of eligible donors.

4.4.5.3 Imputation and Editing Summary for Marital Status

See **Table 9** for a summary of item nonresponse for marital status (recorded in the variable IIMARIT).

Table 9. Marital Status Editing and Imputation Summary

Value of IRMARIT	Assignment of Marital Status	Frequency	Percent
1	From questionnaire	58648	81.72
3	Statistically imputed	30	0.04
9	Legitimate skip (≤14 years old)	13086	18.23

4.4.5.4 Marital Status Recodes

Two additional variables were created from the imputation-revised marital status variable (IRMARIT). MARISTAT had three levels (married, not married, or legitimate skip), and NOTMAR had three levels (never married, divorced/separated or widowed, or married/legitimate skip).

4.4.6 Hispanic-Origin Group

4.4.6.1 Edited Hispanic-Origin Group (EDHOGRP)

The Hispanic-origin group variables divide respondents of Hispanic origin into finer categories. EDHOGRP, the base variable for creating an imputation-revised Hispanic-origin group variable, was created using EDQD04 and the imputation-revised Hispanic-origin indicator (IRHOIND) as follows:

EDHOGRP = EDQD04, if IRHOIND = 1 (Hispanic) and EDQD04 is between 1 and 7, else

99 (legitimate skip), if IRHOIND = 2 (not Hispanic), else

missing.

4.4.6.2 Imputation-Revised Hispanic-Origin Group (IRHOGRP3)

As in the 1999 dataset, IRHOGRP3 had seven possible values (Puerto Rican, Mexican, Cuban, Central or South American, Caribbean islander, other Hispanic, and not Hispanic). It was created using an MPMN method similar to the method for IRRACE. Despite the fact that EDHOGRP had seven levels for respondents of Hispanic origin, the predictive mean vector had only three elements: the predicted probability of the interview respondent falling into each of the first three Hispanic-origin group categories (Puerto Rican, Mexican, and Cuban). This was done to make the computation of both predictive means and Mahalanobis distances[13] more feasible.

The PMN method as applied to the Hispanic-origin indicator is explained in detail in the next four sections: setup for model building, computation of predictive means, assignment of imputed values, and constraints on MPMNs.

4.4.6.2.1 Setup for Model Building

All respondents with IRHOIND = 2 were automatically assigned IRHOGRP3 = 99, and were excluded from the item response propensity models, the predictive mean models, and the sets of potential donors. Imputations were conducted separately within the same three age groups as for the other demographic variables.

---

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

An interview respondent was considered an item nonrespondent for Hispanic-origin group if his/her value for EDHOGRP was missing. The analysis weights of the item nonrespondents were then redistributed among the item respondents using an item response propensity model (see **Appendix C** for the more general GEM), and covariates included Census region, imputation-revised race, gender, age, age squared, age cubed, percent Hispanic population, percent non-Hispanic black population, percent of owner-occupied households, the first-order interaction of age and gender, and the first-order interaction of age squared and gender.

4.4.6.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each of the first three Hispanic-origin group categories was modeled within each age group using polytomous logistic regression.[14] The predictors included in the models were Census region, imputation-revised race, gender, age, age squared, age cubed, percent Hispanic population, percent non-Hispanic black population, percent of owner-occupied households, the first-order interaction of age and gender, and the first-order interaction of age squared and gender.

4.4.6.2.3 Assignment of Imputed Values

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

4.4.6.2.4 Constraints on MPMNs

One logical constraint was placed on potential donors for the Hispanic-origin group variable. If a Hispanic respondent did not indicate a Hispanic group, but he/she did indicate a race when given the opportunity to enter a Hispanic group in the "other" category, donors were constrained to have the same value of IRRACE as the recipient.

In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint states that the donor must have lived in the same segment as the recipient. The second likeness constraint states that each of the donor's three predictive means, as described in **Section 4.4.6.2.2**, must have been within 5 percent of each of

[14]See earlier footnote in **Section 4.4.2.2.2** where a reference for polytomous regression is given.

the recipient's three predictive 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 could be found, the constraint on the predictive means was also removed. See **Appendix G** for the numbers of respondents meeting each set of likeness constraints on sets of eligible donors.

4.4.6.3 Imputation and Editing Summary for Hispanic-Origin Group

See **Table 10** for a summary of item nonresponse for Hispanic-origin group. The imputation indicator variable II2HOGR3 contains this information. As was the case with IIRACE and II2RACE, the imputation indicator IIHOGRP3 is included on the master dataset only for comparison with the 1999 imputation indicator. More detail is contained in II2HOGR3. As with IRRACE, a priority rule[15] is used to determine what group a respondent belongs to if he or she gives more than one response. II2HOGR3 records these cases, whereas IIHOGRP3 merely considers these cases as a "response from questionnaire."

Table 10. Hispanic-Origin Group Editing and Imputation Summary

Value of II2HOGR3	Assignment of IRHOGRP3	Frequency	Percent
1	From questionnaire	8431	11.75
2	From alpha-specify response(s)	882	1.23
3	Statistically imputed (unrestricted)	23	0.03
4	Statistically imputed (restricted by IRRACE)	57	0.08
5	Statistically imputed (unrestricted)	96	0.13
9	Legitimate skip (respondent is not Hispanic)	62371	86.91

4.4.6.4 Hispanic-Origin Group Recodes

HISPGRP (called IRHOGRP in 1999) and HISP2 were created by recoding IRHOGRP3. HISPGRP had five levels: Puerto Rican, Mexican, Cuban, other Hispanic (includes Central or South American and Caribbean islander), and not Hispanic. HISP2 also had five levels: Mexican, Puerto Rican, Central or South American, Cuban, and other (includes other Hispanic, Caribbean islander, and not Hispanic).

[15] The priority rule is the same as that used in past years: Mexican, Cuban, Puerto Rican, Central/South American, Other Hispanic. Details are given in **Appendix D**.

4.4.7 Core Education

4.4.7.1 Edited 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; although EDUC contained codes describing the type of nonresponse, EDEDUC was set to missing if no response was given to QD11.

4.4.7.2 Imputation-Revised Highest Grade Completed (IREDUC)

Due to time constraints, a PMN imputation method was not applied to EDEDUC. As in 1999, an imputation-revised version of EDEDUC was created in 2000 using an unweighted sequential hot-deck procedure to impute missing values. The file was serpentine sorted by population density, segment identification number, marital status, age, race/ethnicity (HISPRACE, see **Section 4.4.4**), and a random number. The imputation-revised version of this variable is called IREDUC. For a summary of nonresponse for highest grade completed, see **Table 11**.

Table 11. Highest Grade Completed Nonresponse Summary

Value of IIEDUC	Assignment of IREDUC	Frequency	Percent
1	From questionnaire	71744	99.97
3	Statistically imputed	20	0.03

4.4.7.3 Education Recode

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

As in the 1999 CAI sample, current work status was a noncore demographic variable determined from multiple questions. Instead of a single question asking the respondent to describe his or her "current" work status, several questions were asked regarding the respondent's work situation during the week prior to the interview and whether that week was atypical. The work status questions were asked only of respondents aged 15 or older.

In 2000, for the first time, 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, 2002a), but missing values were not imputed.

5.2 Current Employment Status

Three imputation-revised employment status variables were created for the 2000 NHSDA: EMPSTAT3, EMPSTT3R, and EMPSTATY. The variable EMPSTATY also has a recoded variable associated with it called EMPSTAT4. **Sections 5.2.1.1** and **5.2.1.2** discuss the edited variables JOBSTAT and EDEMP, respectively. The variable JOBSTAT was used in the creation of EDEMP, which was in turn used as a base variable for EMPSTAT3 (**Section 5.2.2.1**). Because the coding scheme used by the Current Population Survey (CPS) is not consistent with EMPSTAT3, a new variable, called EMPSTT3R (**Section 5.2.2.2**), needed to be created that was consistent with the CPS. It also used JOBSTAT as a starting point for imputation, but a modified version of EDEMP, consistent with the CPS coding scheme, was used as the base variable.

Changes were implemented in the subsequent questionnaire year, 2001, making it impossible to create a 2001 employment status variable consistent with EMPSTAT3 or EMPSTT3R. In particular, respondents in the 2001 questionnaire did not have an opportunity to manually input why they did not have a job last week. New variables had to be created in 2001 to reflect this change. These new variables were also created to be consistent with the CPS coding scheme. Mainly to promote consistency among the variables used in 1999, 2000, and 2001, these new variables were also created in 1999 and 2000 by ignoring the "alpha-specify" responses.[16] EMPSTAT4 (**Section 5.2.2.3**) is most similar to one of the variables from 1999 and 2000 (EMPSTT3R) in that it contains the same levels. However, it is merely a recode of another newly created variable, EMPSTATY (**Section 5.2.2.3**). Unlike EMPSTAT3, EMPSTT3R, or EMPSTAT4, EMPSTATY gives respondents aged 15 to 17 meaningful values instead of skip codes. It uses the edited variables JBSTATR (**Section 5.2.1.3**) and EDEMP4 (**Section 5.2.1.4**).

5.2.1 Edited Employment Status Variables

5.2.1.1 JOBSTAT

One of the variables used to summarize the respondent's current work situation (during the week prior to the interview) was the edited variable JOBSTAT. This edited variable combined the information from questions QD26 to QD31SP, the primary employment status questions in the questionnaire. The categories for JOBSTAT are shown in **Exhibit 2.** Many of the JOBSTAT categories were created using the alpha-specify responses to two questions (QD30 and QD31) regarding why the respondent did not work at a job or business during the week before the interview. These responses were coded in a lookup table, details of which are discussed in **Appendix E**. (The two questions, QD30 and QD31, and their listed responses are also given in **Appendix E**.) Details about the creation of JOBSTAT and the mapping of questionnaire responses (including the codes in **Appendix E**) are discussed in a separate document (Kroutil, 2002a).

5.2.1.2 EDEMP

The base variable EDEMP, which was used to create the imputation revised employment-status variable EMPSTAT3, was derived from JOBSTAT in the following manner:

EDEMP = 5 if the respondent is 12 to 17 years old, else

1 (full-time) if JOBSTAT = 1, 104, 105, 106 or 190, or if JOBSTAT = 3, 6, 7, 8, 102, 103, 192, 193, or 199 and the number of hours usually worked per week is 35 or more (based on QD29), else

[16] "Alpha-specify responses" are written responses entered by the interviewer when the listed responses in the given question were judged as inapplicable (see **Appendices D and E**).

Exhibit 2. Categories of JOBSTAT

Code	Employment Situation	Code	Employment Situation
1	Worked at full-time job, past week	191	Has part-time job, reason for not working unknown
2	Worked at part-time job, past week	192	Has job, did not want/need to work past week
3	Has job but out: vacation/sick/temp absence	193	Has job during school year, no further information
4	Has job but out: layoff, looking for work	199	Has job, no further information
5	Has job but out: layoff, not looking for work	201	Volunteer worker
6	Has job but out: waiting to report to new job	202	Does not need to work
7	Has job but out: self-employed, no business past week	203	Does not want to work
8	Has job but out: in school/training	205	Not eligible/not allowed to work
9	No job: unemployed/layoff, looking for work	206	No job: family responsibilities
10	No job: layoff, not looking for work	207	No job: starting/finished school
11	No job: keeping house full time	208	Student/youth, looking for work
12	No job: in school/training	209	No job: substance abuse issues
13	No job: retired	210	No job: criminal record
14	No job: disabled for work	212	No job: literacy, language, learning disabilities, etc.
101	Seasonal worker	213	Not working due to legal issues
102	Not scheduled/temp/on-call worker	214	Starting new business
103	Babysitter	290	Unemployed, no further information
104	Full-time during school year	291	Doesn't/never worked, reason unspecified
105	Part-time during school year	299	Other, not in labor force
106	Missionary/religious worker	900	Work status unclear
190	Has full-time job, reason for not working unknown	Remaining codes in the 900 series have their standard meanings in the NHSDA: Don't know (994), Refused (997), Blank (998), Legitimate skip (999).	

2 (part time) if JOBSTAT = 2 or 191, or if JOBSTAT = 3, 6, 7, 8, 102, 103, 192, 193, or 199 and the number of hours usually worked per week is less than 35 (based on QD29), else

3 (unemployed) if JOBSTAT = 9, 10, or 290, else

4 (other) if JOBSTAT = 4, 5, 11-14, 101, 201-213, 291, or 299, else

6, if JOBSTAT = 3, 6, 7, 8, 102, 103, 192, 193, or 199 and QD29 was missing, else

missing.

When used as a base variable for EMPSTT3R, the levels of EDEMP are slightly revised. As stated earlier, the only difference between EMPSTAT3 and EMPSTT3R is that EMPSTT3R uses the CPS coding scheme, and EMPSTAT3 does not. The revisions to EDEMP when used as a base variable for EMPSTT3R reflect changes required to maintain consistency with the CPS coding scheme. These revisions are summarized in **Section 5.2.2.2**.

5.2.1.3 JBSTATR

The other variable used to summarize the respondent's current work situation was JBSTATR. The differences between JBSTATR and JOBSTAT are explained in detail in Kroutil (2002a.) The main difference between the two is that JBSTATR does not account for the alpha-specify responses to QD30 and QD31. The categories for JOBSTAT are shown in **Exhibit 3**.

5.2.1.4 EDEMP4

The base variable EDEMP4, which was used to create the imputation revised employment-status variable EMPSTATY, was derived from JBSTATR in the following manner:

EDEMP4 = 5 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 the number of hours usually worked per week is 35 or more (based on QD29), else

2 (part time) if JBSTATR = 2 or 191, or if JBSTATR = 3, 6, 7, 8, or 199 and the number of hours usually worked per week is less than 35 (based on QD29), else

Exhibit 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: unemployed/layoff, 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 the NHSDA: Don't know (994), Refused (997), Blank (998), Legitimate skip (999).	
11	No job: keeping house full time		

3 (unemployed) if JBSTATR = 4, 5, 9, or 10, else

4 (other) if JBSTATR = 11-15, 290, or 299, else

6, if JBSTATR = 3, 6, 7, 8, 102, or 199 and QD29 was missing, else

missing.

5.2.2 Imputation-Revised Employment Status (EMPSTAT3)

5.2.2.1 EMPSTAT3

Missing values in the edited employment status variable EDEMP were replaced with imputed values using an unweighted sequential hot-deck procedure. This procedure is described in greater detail in **Appendix A**. The imputation procedure for EMPSTAT3 was not changed from the 1999 CAI NHSDA.

Respondents aged 12 to 17 were automatically assigned EMPSTAT3 values of 5 and were separated from the rest of the file prior to imputation, so they could not act as donors for item nonrespondents older than 17. These cases were added back into the file following imputation. The file used for imputation was serpentine sorted by highest grade completed, race/ethnicity, gender, age, and a random number. Missing values were imputed as follows:

- If a record had a missing employment status and had an EDEMP code of "6," an imputed response was set equal to that of the previous record on the sorted file who reported either working full time or part time. Recall that these respondents were known to be employed based on their JOBSTAT value; thus, their current employment status could not be "unemployed" or "other."

- If a record had a missing employment status and did not have an EDEMP code of "6," an imputed response was set equal to that of the previous record in the file.

For a summary of item nonresponse for EMPSTAT3, see **Table 12**.

Table 12. EMPSTAT3 Editing and Imputation Summary

Value of IIEMPST3	Assignment of EMPSTAT3	Frequency	Percent
1	From questionnaire	45684	63.66
3	Statistically imputed	363	0.51
4	12 to 17 years old	25717	35.84

5.2.2.2 EMPSTT3R (Revised EMPSTAT3)

The only difference between EMPSTAT3 and EMPSTT3R is in the treatment of JOBSTAT levels 4, 5, and 290. It was decided that levels 4 and 5 should be mapped to "unemployed" and level 290 should be mapped to "other" to be consistent with the CPS definition of "unemployed." No corresponding imputation indicator was created for EMPSTT3R because its values would be exactly the same as those for IIEMPST3.

5.2.2.3 EMPSTATY and EMPSTAT4

The variable EMPSTATY used EDEMP4 as its base variable instead of EDEMP. The imputation method was identical to that of EMPSTAT3 and EMPSTT3R, except that only respondents aged 12 to 14 were taken out of the dataset and assigned a skip code of 5. For a summary of item nonresponse for EMPSTATY, see **Table 13**. EMPSTAT4 is a recode of EMPSTATY, in that the responses for 15 to 17 year olds are recoded to a skip code of 5. In this way, EMPSTAT4 is more similar to both EMPSTAT3 and EMPSTT3R rather than EMPSTATY.

Table 13. EMPSTATY Editing and Imputation Summary

Value of IIEMPSTY	Assignment of EMPSTATY	Frequency	Percent
1	From questionnaire	58562	81.60
3	Statistically imputed	116	0.16
4	12 to 17 years old	13086	18.23

6. CAI Drug Imputations

6.1 Introduction

Major changes were introduced in the imputation procedures for the drug use variables in the CAI sample of the 1999 NHSDA. In particular, a new imputation methodology (i.e., predictive mean neighborhood [PMN]) was developed specifically for the NHSDA. This methodology is a combination of weighted regression and nearest neighbor hot-deck imputation, where the hot deck is random whenever possible. Its application to the drug use variables in the 2000 NHSDA was expanded slightly from 1999, as is explained in the following sections.[17]

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 either 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 be edited to past year usage, where past month versus past-year-but-not-past-month use is determined by imputation. The aforementioned editing processes are summarized by Kroutil (2002a).

The models for these imputations, which are described in detail in the following sections, were either binomial or multinomial weighted logistic models, 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 is constructed from a univariate predicted mean, the assignment of imputed values may be either univariate or multivariate. The members of the neighborhood were restricted to satisfy two types of constraints: "logical constraints" and "likeness constraints." Constraints that make the imputed values consistent with preexisting values of other variables are called logical

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

constraints and are required for the candidate donor to be a member of the neighborhood. Likeness constraints are implemented to make donors and recipients as much alike as possible. Although logical constraints cannot be loosened, likeness constraints can be loosened if they force the donor pool to be too sparse. Details of these imputation procedures are given in **Appendix C**.

Because drug use is 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 year olds, 18 to 25 year olds, and persons 26 years of age or older.[18]

Although statistical imputation of the drug use variables could not proceed separately within each State due to insufficient pools of donors, information about the State of residence of each respondent is incorporated in the modeling and hot-deck steps in the CAI sample. States were classified into three drug usage categories: 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 1999 CAI NHSDA, the 2000 NHSDA has different drugs and drug use measures than are found in pre-1999 NHSDAs. **Exhibit 4** 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, 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

---

[18] Modeling was done separately within each of the three age groups regardless of the response variable.

Exhibit 4. Drugs and Drug Use Measures, Univariate Versus Multivariate Imputation

Drug	Drug Use Measure						
	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[1]	✓✓	✕✕		✕✕		✓✕	
Cigars	✓✓	✕		✕		✓	
Pipes	✓✓	✓					
Alcohol	✓✓	✕	✕	✕	✕	✓	
Inhalants	✓✓	✕	✕	✕		✓	
Marijuana	✓✓	✕	✕	✕		✓	
Hallucinogens[2]	✓✓	✕✕	✕	✕		✓	
Pain Relievers	✓✓	✕	✕			✓	
Tranquilizers	✓✓	✕	✕			✓	
Stimulants[3]	✓✓	✕✕	✕			✓✕	
Sedatives	✓✓	✕	✕			✓	
Cocaine and Crack	✓✓	✕✕	✕✕	✕✕		✓✕	
Heroin	✓✓	✕	✕	✕		✓	

✓ Univariate neighborhood; univariate assignment of imputed values.

✓✓ Multivariate neighborhood across all lifetime drug use variables; multivariate assignment of imputed 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 (e.g., see **Sections 6.4.1.7.1**, **6.4.1.7.2**, **6.4.1.7.3**, and **6.4.2.7**).

✓✕ Univariate neighborhood and multivariate assignment of imputed values (see **Sections 6.5.1.7.1**, **6.5.1.7.2**, and **6.5.1.7.3**).

[1] Includes chewing tobacco and snuff.
[2] Includes LSD and PCP.
[3] Includes methamphetamines.

sequence could be used as covariates for models fitted later in the sequence. Because the gate questions are the basis for all subsequent drug data, the imputation of missing values for lifetime drug use for all drugs must precede imputations of all other drug use measures. These lifetime use indicators are temporary in the sense that they are manifested within the drug recency and frequency-of-use variables, but are not delivered themselves. The hierarchy of models for drugs for the lifetime usage models is discussed in **Section 6.3**.

Once all the lifetime usage indicators had been determined, the imputations of the remaining measures could proceed. As indicated in **Exhibit 4**, a multivariate imputation was implemented across the measures within each drug for recency of use, 12-month frequency of use, 30-day frequency of use, and binge drink frequency (alcohol only). For a given drug, recency of use 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. Finally, age at first use must be consistent (in a number of ways) with the other measures (see **Section 6.5**). Hence, age at first use was imputed after the imputation for the other measures was completed.[19] The following sections describe the imputation procedures for each drug use measure.

6.3 Imputing Lifetime Drug Use Indicators

As with the 1999 CAI NHSDA, the 2000 NHSDA implemented automatic routing through the questionnaire. Using a series of gate questions, the instrument asked the respondent 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 skip 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 key to the range of responses available for subsequent questions in each module.

6.3.1 Hierarchy of Drugs

The first step in the imputation of lifetime indicators was to determine the order in which the drugs would be modeled (i.e., the "drug hierarchy" discussed in detail in **Appendix C**). For a particular drug, it was expected that indications of lifetime use of other drugs would be strong predictors of lifetime use of that drug. Hence, drugs expected to be highly correlated with the lifetime use of other drugs were placed later in the sequence. It is important to note that the lifetime usage indicators, when used as predictors, were only provisional because the final

[19] For cigarettes, both age at first use and age at first daily use had to be 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.

imputation of lifetime usage indicators was not implemented until the lifetime usage modeling was completed for all drugs. The order in which the lifetime indicators of use were imputed is shown in **Exhibit 5**.

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 year olds, 18 to 25 year olds, and respondents 26 years of age or older. For an individual to be considered a lifetime-use item respondent, he or she must have 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, hallucinogens other than LSD and PCP, analgesics, tranquilizers, methamphetamines, stimulants other than methamphetamines, and sedatives. Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. The predicted probability *P* (survey respondent is an item respondent | respondent is a lifetime user) was determined for each item respondent from this model, the inverse of which was multiplied by the respondent's weight. Due to the fact that item respondents were defined across all drugs, this adjustment was only computed once per age group and then used in the modeling of lifetime use for all drugs. The item response propensity model is a special case of the generalized exponential model (GEM), which is described in greater detail in **Appendix B**.

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, 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 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; and (10) any other aerosol spray. If the response to any of

Exhibit 5. Lifetime Indication of Use ("Gate") Questions for CAI (in Order of Imputation)

Drug	Question(s)
Cigarettes	CG01
Smokeless Tobacco[1]	CG17, CG25
Cigars	CG34
Pipes	CG42
Alcohol	AL01
Inhalants	IN01a, IN01b, IN01c, IN01d, IN01e, IN01f, IN01g, IN01h, IN01i, IN01j, IN01l
Marijuana	MJ01
Hallucinogens[2]	LS01a, LS01b, LS01c, LS01d, LS01e, LS01f, LS01h
Pain Relievers	PR01, PR02, PR03, PR04, PR05
Tranquilizers	TR01, TR02, TR03, TR04, TR05
Stimulants[3]	ST01, ST02, ST03, ST04, ST05
Sedatives	SV01, SV02, SV03, SV04, SV05
Cocaine	CC01
Crack	CK01
Heroin	HE01

[1] Includes chewing tobacco and snuff.
[2] Includes LSD and PCP.
[3] Includes methamphetamines.

these questions was "yes," the respondent was deemed a lifetime user of inhalants, even if some of the other responses to the gate questions in the inhalants module were unanswered. Similarly, composite lifetime indications of use were formed for hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, and smokeless tobacco. To be considered a nonrespondent of a drug module with multiple gate questions, the respondent had to answer "no" to all of the gate questions. If none of the gate questions in a drug module was answered affirmatively, but some 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 was used to determined the parameter estimates. Because the interest was only in the estimation of the predicted mean, and not in the parameter estimates (by themselves) or their standard errors, no model selection was attempted. The predictors in each model included continuous age, age squared, age cubed, race/ethnicity, gender, lifetime use of drugs already imputed, Census region, population density, a three-level State rank variable (incorporating the proportion of lifetime users of the drug of interest in the respondent's State of residence), and first-order interactions of age, and gender. For age groups 18 years of age or older, the variables for marital status, education, and employment status were also included. For a complete summary of the lifetime use imputation models, see **Appendix F**.

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

Using the parameters 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. 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[20] 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 lifetime use.[21] In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and 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 zero or 1. The range of values for delta across various predicted probabilities is given in **Exhibit 6**. 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.

---

[20] "Delta" refers to the value that defines the neighborhood of donors that are "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 be less than delta. See **Appendix C** for more details.

[21] The probability of past month use was used to define univariate neighborhoods even when it was known that the respondent was not a past month user. More details are provided on this matter later in this section.

Exhibit 6. Values of Delta for Various Predicted Probabilities of Lifetime Use

Predicted Probability (*p*)	Delta
$p \leq 0.5$	$0.05*p$
$p > 0.50$	$0.05*(1-p)$

6.3.5 Assignment of Provisional Imputed Values

Separate assignment of provisional values were performed within each of the three age groups, subject to the constraints described in the next section. 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. 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 could also be considered a likeness constraint, which could be loosened by enlarging delta. This was never done, however, with the lifetime usage indicators.

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 carried out 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 particular drugs sometimes required the assignment step to be multivariate. Drugs for which a multivariate assignment was necessary are discussed below.

6.3.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

Many respondents who indicated lifetime use of smokeless tobacco seemed to be confused regarding the difference between chewing tobacco ("chew") and snuff, as was demonstrated by their responses to questions regarding specific brands. For example, many respondents who indicated use of chewing tobacco entered a snuff brand, such as "Copenhagen™," when asked about the specific brand of chew they used. As a result, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted, rather than individual models for chew and snuff. The 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 because chew and snuff were assigned values independently, 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 CAI questionnaire, separate models were fit 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 given in **Exhibit 6**, based on both the cocaine- and crack-predicted probabilities of lifetime use. An item respondent was eligible to be 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. 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 be possible for a respondent to be 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, and Other Hallucinogens) and Stimulants (Methamphetamines and Other Stimulants)

The modules for both hallucinogens and stimulants included multiple gate questions (called subgate questions), and some of the substances referred to in the subgate questions were of interest in their own right. For hallucinogens, there was interest in the usage of LSD and PCP; for stimulants, there was interest in the usage of methamphetamines. Predicted probabilities were calculated for the larger groups of substances known as hallucinogens and stimulants, 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 except the ones of special interest. In the final assignment step, lifetime usage indicators were assigned for LSD, PCP, and "other" for hallucinogens, and for methamphetamines and "other" for stimulants. The final lifetime usage indicators for hallucinogens and stimulants 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 and PCP. 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' subgates 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 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 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 all hallucinogens was created by combining the lifetime usage indicators for the three subgroups.

6.3.7.3.2 Stimulants

The procedure for stimulants followed the same pattern used for hallucinogens. A lifetime usage indicator for "other stimulants" was created using information from all the stimulants' subgate questions except methamphetamines. As with hallucinogens, a respondent's other stimulants' lifetime usage indicator was only missing if the subgate questions other than methamphetamines were all unanswered, or were a combination of unanswered questions and "no" responses. Using the neighborhood created from the stimulants' predicted probability of lifetime use, the missing value(s) for methamphetamines and/or other stimulants was (were) replaced with the value(s) from a donor within this neighborhood. For individuals missing a lifetime usage indicator for either methamphetamines or other stimulants but not both, only the missing value was replaced. For individuals missing both of these lifetime usage indicators, the missing values were replaced by values from the same donor. As with smokeless tobacco, the subcategories for stimulants were assigned values separately, making logical constraints unnecessary. As a final step, a lifetime usage indicator for all stimulants was created by combining the lifetime usage indicators for the two subgroups.

6.3.8 Multivariate Imputation for Lifetime Drug Use

Section 6.3.2 summarizes how all of the respondents in the 2000 NHSDA were separated into item respondents and item nonrespondents for the lifetime drug variables. The sections following **Section 6.3.2** 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 **Exhibit 4**, 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 year olds, 18 to 25 year olds, and respondents 26 years of age or older. As indicated in earlier sections, a respondent was eligible to be 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.

As with the univariate imputations discussed in **Section 6.3.6**, no logical constraints were utilized in the multivariate imputation of lifetime use. 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 only missing one lifetime use indicator were imputed using UPMN. Respondents missing more than one lifetime use indicator were imputed using MPMN.

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 predicted mean vector "close to" (i.e., within the delta distance of) the recipient's elements of the predicted mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predicted mean vector. The elements of the predicted 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 only required to have the same values for lifetime indicators within the same or related drug modules. The number of respondents for whom donors could be 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 predicted mean vector contained elements for each lifetime drug use measure. However, only a portion of the full predicted mean vector was used; that is, 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, the provisional UPMN values were considered final. Otherwise, an MPMN imputation was employed. The Mahalanobis distance was then calculated using only the portion of the predicted mean vector associated with the given missingness pattern.[22] If no donors were available that 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**.

---

[22] See **Appendix C** for a definition of Mahalanobis distance.

6.4 Imputation-Revised Drug Recency, 12-Month Frequency of Use, and 30-Day Frequency of Use Variables Created for Completed Cases

In the 2000 NHSDA, the drug use measures' 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[23] 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 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) was employed to determine final imputed values using the predicted values from these models. Separate multivariate imputations were conducted for each drug. If no donor could be found using the MPMN technique, even after loosening likeness constraints, UPMN values were used as final imputed values.

The implementation of the PMN methodology requires 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.

6.4.1 Recency of Use

6.4.1.1 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 arise 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.) The cigarettes' recency was modeled first, and the predicted probability of past month use was used to determine provisional values[24]

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

[24] Although the final imputation was multivariate across drug measures, provisional versions of the drug recencies were created using the UPMN methodology described in **Appendix C**.

for the cigarette frequency models. The final imputation-revised cigarette recency was used in the models for cigars, smokeless tobacco, pipes, and alcohol recency of use. Once the multivariate imputations for the tobacco products were complete, the alcohol imputations were conducted. Unlike the sequences used for lifetime usage and age at first use, marijuana followed alcohol and inhalants followed marijuana, rather than vice versa because marijuana recencies were needed for subsequent models. After inhalants, the sequence was exactly the same as the sequence used for lifetime usage.

6.4.1.2 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 subset down 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. Because item respondents were defined at the drug level, these adjustments were made separately for each drug (and within the three age groups). Adjustments were made using an item response propensity model (see **Appendix B** for the more general GEM), and covariates included a continuous age; age squared; gender; race; first-order interactions of age, gender, and race; marital status; education; employment status[25]; Census region; an MSA[26] indicator; imputation-revised cigarette, alcohol, and marijuana recencies (where applicable); and lifetime indicators of usage of drugs other than cigarettes, alcohol, and marijuana. 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.[27]

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

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

[27] In a handful of cases (e.g., heroin, any age group), it was necessary to abandon the State rank variable due to the small number of users and the convergence difficulties that resulted when the State rank variable was in the model.

6.4.1.3 Sequential Model Building

Using the adjusted weights, the probability of selecting each recency-of-use category was modeled within each age group using polytomous logistic regression. The predictors included in the models were age; age squared; gender; race; first-order interactions of age, gender, and race; marital status; education; employment status[28]; Census region; an MSA indicator; State rank; imputation-revised cigarette, alcohol, and marijuana recencies (where applicable); and lifetime indicators of usage of drugs other than cigarettes, alcohol, and marijuana. Because interest was only in the estimation of the predicted mean, and not in the parameter estimates (by themselves) or their standard errors, no model selection was attempted. For a summary of the variables included in each drug model, see **Appendix F**.

6.4.1.4 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 be 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 polytomous logistic model. 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 polytomous logistic regression model. Because only a single predicted mean was used to determine the neighborhood when determining provisional values, not all the predicted probabilities from the model were used.[29] 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[30] of the

---

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

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

[30] "Delta" refers to the value that defines the neighborhood of donors that are "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 be less than delta. See **Appendix C** for more details.

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.[31] In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and 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 zero 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.

6.4.1.5 Assignment of Provisional Imputed Values

Separate assignments of provisional values were performed within each of the three age groups, subject to the constraints described in the next section. The final recency-of-use imputations were multivariate across drug measures and are further described in **Section 6.4.5**.

6.4.1.6 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. As with all other drug use measures, neighborhoods for recency of use were restricted so that candidate donors and recipients would be 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 could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. As previously stated, if no donors could be found in the delta as defined in **Section 6.4.1.4**, the neighborhood was abandoned, and the donor with the predicted mean closest to the recipient was chosen.[32] 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 to 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

---

[31] The probability of past month use was used to define univariate neighborhoods even when it was known that the respondent was not a past month user. More details are provided on this matter later in this section.

[32] Although using neighborhoods is important for 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 by the time these imputations were implemented.

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. **Appendix G** gives a summary of how many respondents had values imputed under various constraints.

Logical constraints were placed on the neighborhoods in those 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, and the restrictions on possible donors that did not involve an interview date, are given in **Exhibit 7**. As indicated in the exhibit, an additional logical constraint was applied only to tobacco products: If the respondent's age at first use was within 2 years of his or her current age, it would be impossible for a respondent to have last used the substance more than 3 years ago. Hence, under these circumstances, the donors were limited to have used within the past 3 years. Such a logical constraint would not be useful for nontobacco products because the recency categories for lifetime use but not past 3 year use and for past 3 year use but not past year use were combined into a single category for lifetime use but not past year use. Additional logical constraints, not listed in Exhibit 7, limited the possible recencies that could be assigned based on the respondent's current age, the time between the interview date and the birth date, the time between the interview date and the month of first use, and any nonmissing frequency-of-use information. A complete list of missingness patterns across recency and frequency of use (including patterns with general recency categories), and the logical constraints that correspond to those missingness patterns, is given in **Appendix H**. See **Section 6.4.5** for a discussion of the multivariate imputation of recency and frequency of use.

Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. Those cases are discussed in detail in the next section (**Section 6.4.1.7**).

Exhibit 7. Logical Constraints on Univariate Predictive Mean Neighborhoods (Not Involving Interview Date) When a General Recency Category Was Given

General Recency Cate-gory	Combination of Specific Recency Categories (Tobacco)	Combination of Specific Recency Categories (Nontobacco)	Logical Constraints (Tobacco)	Logical Constraints (Non-tobacco)
Lifetime	1. Lifetime, not past 3 years 2. Past 3 years, not past year 3. Past year, not past month 4. Past month	1. Lifetime, not past year 2. Past year, not past month 3. Past month	If age at first use was within 2 years of current age, donors must have used in the past 3 years	No constraints
Lifetime, Not Past Year	1. Lifetime, not past 3 years 2. Past 3 years, not past year	N/A (for nontobacco, this is a specific recency category)	Donors must not have used in the past year	N/A
Lifetime, Not Past Month	1. Lifetime, not past 3 years 2. Past 3 years, not past year 3. Past year, not past month	N/A	1. Donors must not have used in the past month 2. If age at first use was within 2 years of current age, donors must have used in the past 3 years	N/A
Past Year	1. Past year, not past month 2. Past month	1. Past year, not past month 2. Past month	Donors must be past year users	Donors must be past year users

6.4.1.7 Multivariate Assignments

Although the methodology for determining the neighborhood was univariate in terms of the predicted probability of past month use, peculiarities associated with particular drugs sometimes required the assignment step to be multivariate. Drugs for which a multivariate assignment was necessary are discussed below.

6.4.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

For reasons discussed in **Section 6.3.7.1**, one model for smokeless tobacco (a combination of the chew and snuff responses) was fit 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. 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[33] for either chew or snuff but not both, only the missing specific recency-of-use values were replaced. However, for individuals missing recency-of-use information for both chew and snuff (given that the respondent was known or was imputed to be 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 snuff and chew.

Unlike recency of use, separate models for snuff and chew 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 6.4.3.3**, 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 can be obtained from the recency models. Because the 30-day frequencies for snuff and chew could not be combined, recency-of-use models were built for snuff and chewing tobacco separately, where the response was past month use versus not past month use. (This was in addition to the regular recency-of-use model that was built for smokeless tobacco.) See **Section 6.4.3.3** for more details. The covariates used in the models are the given in **Appendix F**.

6.4.1.7.2 Cocaine and Crack

Even though cocaine and crack are in distinct modules in the CAI questionnaire, a recency model was only fit for cocaine. Crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. As with the other drugs, the neighborhood was defined in terms of delta, 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 of cocaine. In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. 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 information required donors who were also crack users. Moreover, if the cocaine recency was not missing, the donated crack recency could not be more recent than

[33] For respondents missing all of their recency information, the only known information is 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 **Exhibit 7**), and specific recency values needed to be imputed.

the preexisting cocaine recency. Once the neighborhood was defined, missing specific recency-of-use categories for crack and/or cocaine were replaced with the values from a donor within this neighborhood. For individuals missing specific recency-of-use categories for only crack, but not both crack and cocaine, only the missing categories for crack were replaced. However, for individuals missing both crack and cocaine, all missing recency-of-use information was replaced by values from the same donor.

6.4.1.7.3 Hallucinogens (LSD, PCP, and Other Hallucinogens) and Stimulants (Methamphetamines and Other Stimulants)

As stated in **Section 6.3.7.3**, the modules for hallucinogens and stimulants included subgate questions referring to substances that were of interest in their own right. For hallucinogens, there was interest in the usage of LSD and PCP; for stimulants, there was interest in the usage of methamphetamines. Recency-of-use information for both hallucinogens and stimulants was used in subsequent models; LSD, PCP, and methamphetamines' recencies of use were not used. Hence, obtaining provisional values for the recency of use of the substances corresponding to the subgate questions was less crucial. The imputed values for these substances were still retained in case final values could not be determined using the MPMN technique.

Predicted recency probabilities were calculated for the larger groups of substances known as hallucinogens and stimulants, and these probabilities were used to determine neighborhoods for each group of drugs. As with smokeless tobacco, use or nonuse of LSD, PCP, and methamphetamines was considered given (employing the lifetime usage imputations).

Hallucinogens. Using the neighborhood created from the predicted probability of past month use of hallucinogens, missing specific recency categories for LSD and/or PCP and/or hallucinogens as a whole were replaced with the specific recency categories from a single donor. LSD users and PCP users with incomplete recency information were constrained to have donors who were LSD users and PCP users, respectively. Moreover, donors were constrained so that a preexisting LSD or PCP recency could not be more recent than a donated hallucinogens recency; conversely, a preexisting hallucinogens recency could not be less recent than donated LSD or PCP recency. For individuals missing recency information for either LSD and/or PCP and/or hallucinogens as a whole, only the missing value(s) was (were) replaced. For individuals missing recency information on two or more of these substances, the missing categories were replaced by values from the same donor.

Stimulants. A similar procedure was followed for the stimulants module. Using the neighborhood created from the stimulants' predicted probability of lifetime use, missing specific recency-of-use categories for methamphetamines and/or stimulants as a whole were replaced with the specific recency categories from a single donor within this neighborhood. Methamphetamine users with incomplete recency information were constrained to have donors who were also methamphetamine users. Moreover, donors were constrained so that a preexisting methamphetamine recency could not be more recent than a donated stimulant recency, and conversely, a preexisting stimulant recency could not be less recent than donated methamphetamine recency. For individuals missing recency information for methamphetamines and/or hallucinogens 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.

6.4.2 12-Month Frequency of Use

6.4.2.1 Hierarchy of Drugs

The modeling of 12-month frequency followed that of recency of use for each drug. Across drugs, the sequence was exactly the same as that 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 **Exhibit 4**).

6.4.2.2 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 response propensity adjustment modeling included age, race, gender, Census region, an MSA indicator, and (where available) recencies of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives as predictors.[34]

6.4.2.3 Model Building

As was apparent from the previous section, only past year users of the drug of interest were used to build the 12-month frequency-of-use model. The (untransformed) response variable of interest in the 12-month frequency-of-use models for most respondents was the proportion of the days in a full year (365.25) on which a respondent used a particular drug. For example, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100 / 365.25. Some respondents, however, started using the drug within the past year. If they responded to the month at first use question, the difference between the month at first use and the date of the interview indicated the total time period during which they could have used.[35] If the date of the interview was July 10th, for example, and the month of first use was March, the maximum period during which the respondent could have used is the number of days between March 1st and July 10th, or 101. Thus, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/101 instead of 100 / 365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

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

where Y_i is the observed 12-month frequency for respondent i and N is the total number of days in the year that the respondent could have used the substance. This transformation is nearly equivalent to the standard logit transformation:

$$Y_i = \ln[P_i/(1 - P_i)],$$

[34] 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 vs. past year but not past month use) was always defined.

[35] If a respondent initiated use in the past year (according to his/her age at first use response), but did not answer the month at first use question, the maximum period the respondent could have used was assumed to be 365.25 because no other information is available.

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

Because the 12-month frequency models were limited to past year users, only two recency categories could result: past month use and past year but not past month use.[37] 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 were used if available. For drugs for which the recency of use was not yet modeled, the lifetime indication of use served as a surrogate for the recency-of-use indicators. 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.[38] Thus, the models included age; age squared; age cubed; gender; race; State rank (based on past month prevalence of the drug); marital status; employment; educational level[39]; Census region; an MSA indicator; (where available) the provisional recencies of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; as well as first-order interactions of age, gender, and race.[40] Because interest focused only on the estimation of the predicted mean, and not on the parameter estimates (by themselves) or their standard errors, no model selection was attempted. Predicted 12-month frequencies of use were defined by back-

---

[36] If the respondent was a daily user of the substance, then $\log[(Y + 0.5) / (N - Y + 0.5] \simeq \log[N + 0.5 / 0.5]$, so that it is defined for all respondents. See Cox and Snell (1989) for a discussion of the empirical logistic transformation.

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

[38] As with the recency-of-use models, for a handful of cases the State rank variable could not be included in the model. Usually, but not always, the age group/drug combination that had problems was the same for recency of use and 12-month frequency of use.

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

[40] The covariate based on the recency-of-use variable for the same drug as the one being modeled was a single dummy variable indicating past month use or nonuse, as described previously. The covariates based on recency-of-use variables that corresponded to drugs other than the one being modeled were defined by a series of dummy variables reflecting the different recency categories. Lifetime usage indicators were used instead of the recency-of-use indicators when recency of use was not available.

transforming the resulting predicted values. For a complete summary of the 12-month frequency-of-use models, see **Appendix F**.

The predicted mean that results from the 12-month frequency-of-use model is 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 can be treated as a probability, which in turn could be multiplied by the probability of past year use to make the predicted mean conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it is not known whether they are past year users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

6.4.2.4 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. 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[41] 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 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 a tighter delta for predicted proportions close to zero 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.[42]

---

[41] "Delta" refers to the value that defines the neighborhood of donors "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 be less than delta. See **Appendix C** for more details.

[42] Although using neighborhoods is important for 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 by the time these imputations were implemented.

6.4.2.5 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, or for "pills" for the reasons given in the previous section. The final 12-month frequency-of-use imputations were multivariate across drug measures and are further described in **Section 6.4.5**.

6.4.2.6 Constraints on Univariate Predictive Mean Neighborhoods

An obvious logical constraint for 12-month frequency of use was that all donors must be past year users, whether that past year use is reported or (provisionally) imputed. Other logical constraints involved the interview date, month of first use, birthday, recency of use, and 30-day frequency of use. A complete listing of missingness patterns, and the logical constraints associated with those missingness patterns, is given in **Appendix H**. See **Section 6.4.5** 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 vs. past year not past month), whether that recency of use was reported or imputed.[43] 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; (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, leading to a multivariate assignment of imputed values. Those cases are discussed in detail in the next section.

6.4.2.7 Multivariate Assignments

Although the methodology for determining the neighborhood was univariate in terms of the predicted proportion of the year used (or maximum possible period within the year used), peculiarities associated with particular drugs sometimes required the assignment step to be multivariate. Drugs for which a multivariate assignment was necessary are discussed below.

Even though cocaine and crack are in distinct modules in the CAI questionnaire, a 12-month frequency-of-use model was only fit for cocaine. Donors for crack and cocaine were obtained using a single neighborhood, which was defined in the same manner as for the other drugs.[44] As with recency of use, use or nonuse of crack was considered given (using information from the lifetime imputations). In the same manner as for the drugs where univariate assignments were required, recipients and donors were required to have the same cocaine recency of use, whether that recency of use was reported or imputed. In addition, donors and recipients were also required to have the same crack recency of use if the recipient used crack in the past year.[45] Both of these constraints were applied whether the recipient was missing the 12-month

---

[43] Because all respondents in the 12-month frequency of use imputation were past year users by definition, this meant that 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 fit that specific recency category.

[44] Delta was set so that donors required a predicted proportion within 5 percent of that of the item nonrespondent. If insufficient donors were available within 5 percent, the neighborhoods were dropped and the item respondent with the closest predicted mean was chosen.

[45] If, in the original data, the respondent was missing both the recency and 12-month frequency, but the provisional imputed value for recency of use was lifetime but not past year use, no imputation was required for 12-month frequency. Such a respondent, however, might be imputed to one of the past year use categories with a corresponding 12-month frequency in the final MPMN imputation.

frequency for only cocaine, only crack, or both. 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 crack and cocaine. If the 12-month frequencies for both crack and cocaine were missing, this 12-month frequency product for crack could not be greater than that of cocaine. If only the crack 12-month frequency was missing, the donated 12-month frequency product for crack could not be greater than the observed cocaine 12-month frequency; conversely, if only the cocaine 12-month frequency was missing, the donated 12-month frequency product for cocaine could not be less than the observed crack 12-month frequency. Finally, if the observed 12-month frequency for cocaine was 1, and the 12-month frequency for crack was missing but the respondent was a past year user of crack, naturally the 12-month frequency for crack should be 1.

Once the neighborhood was defined, the missing 12-month frequency was determined by taking the product of the donated proportion(s) and the recipient's maximum number of possible days used for crack and/or cocaine. For individuals missing a 12-month frequency for only crack, but not both crack and cocaine, only the missing value was replaced. However, for individuals missing both crack and cocaine, both 12-month frequencies were replaced by values from the same donor.

6.4.3 30-Day Frequency of Use

6.4.3.1 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 only for a subset of the drugs (see **Exhibit 4**).

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

The file was first subset down to the eligible population: past month users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the recency and 12-month frequency imputations. To be an item respondent, the individual had to provide 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. Predictors for the response propensity models included age; race; gender; Census region;

an MSA indicator; imputation-revised recencies of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; and the provisional 12-month frequency for the drug of interest (where applicable).

6.4.3.3 Model Building

As was 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 (untransformed) response variable of interest in the 30-day frequency-of-use models for most drugs 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:

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

where Y_i was the observed 30-day frequency for respondent i and N was the total number of days in the year that the respondent could have used the substance. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i = \ln[P_i/(1 - P_i)],$$

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.[46] Using the adjusted weights, a linear univariate regression model was then fit 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.[47] Hence, provisional recency of use for the drug of interest could not be included in the 30-day frequency-of-use model. However, imputation-revised recency of use of other drugs could be included. For drugs where

---

[46] If the respondent was a daily user of the substance, then $\log[(Y + 0.5) / (N - Y + 0.5] \approx \log[N + 0.5 / 0.5]$, so that it is defined for all respondents. (See Cox and Snell, 1989, for a discussion of the empirical logistic transformation.)

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

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 age; age squared; age cubed; gender; race; marital status; employment; educational level[48]; Census region; an MSA indicator; imputation-revised recency-of-use values for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; the provisional 12-month frequency of use for the drug of interest (where applicable); and the first-order interactions of age, gender, and race. Because interest was only in the estimation of the predicted mean, and not in the parameter estimates (by themselves) 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 comes out of the 30-day frequency-of-use model is 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 can be treated as a probability, which in turn could be multiplied by the probability of past month use to make the predicted means conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it is not known whether they are 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, snuff, and chewing tobacco, 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 which 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 snuff and chewing tobacco could not be combined into a single

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

value for smokeless tobacco. One could not know if *x* days using snuff overlapped with the *y* days using chewing tobacco. Hence, separate models were fit for snuff and chewing tobacco.

6.4.3.4 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 fit 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).

6.4.3.5 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.

6.4.3.6 Constraints on Univariate Predictive Mean Neighborhoods (Alcohol Only)

An obvious logical constraint is that all donors had to be past month users, whether that past month usage was reported or (provisionally) imputed. In addition, the donated 30-day frequency was required to be 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 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 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; (2) donors and recipients were no longer required to be from States with similar usage levels.

Although a multivariate assignment was necessary in the final imputation for crack and cocaine, no multivariate assignment of provisional imputed values was required for the 30-day frequency.

6.4.4 30-Day Binge Drinking Frequency

In addition to the 30-day frequency of use, an additional frequency variable was defined for alcohol: the number of days in the past month during which the respondent had five or more drinks, or the 30-day binge drinking frequency, also known as DR5DAY. 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[49] 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 same weight adjustment was used in the modeling of the 30-day binge drinking frequency as was used for the 30-day frequency model.

The (untransformed) response variable of interest in the 30-day binge drinking frequency models for most drugs 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 frequency of use, the following empirical logit transformation was computed for all respondents:

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

where Y_i was the observed 12-month frequency for respondent i and N was the total number of days in the year that the respondent could have used the substance. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i = \ln[P_i/(1 - P_i)],$$

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

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

The predicted means from this model were used solely in the multivariate predicted mean vector used in the final MPMN imputation. No UPMN step was taken, and no provisional imputed values were determined.

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

Sections 6.4.1, **6.4.2**, and **6.4.3** summarize how the set of lifetime drug users in the sample of the 2000 NHSDA 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 **Exhibit 4**, 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 year olds, 18 to 25 year olds, and respondents 26 years of age or older. As indicated in earlier sections, a respondent was eligible to be 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.

The logical constraints required in the univariate imputations discussed in **Sections 6.4.1**, **6.4.2**, and **6.4.3** were also required in the multivariate imputations. 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

---

[50] 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 logistic transformation.)

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. 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 given in **Appendix H** (see Section H.2).

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 predicted mean vector "close to" (i.e., within the delta distance) the recipient's elements of the predicted mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predicted mean vector. Finally, for drug modules with multiple substances, 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. The number of respondents for whom donors could be 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 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; then (3) donors and recipients were no longer required to be from States with similar usage levels.

The full predicted 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 predicted mean vector that could potentially result are given in **Exhibit 8**. It is important to note that not all drugs contained all the elements given. **Exhibit 9** shows the full predicted mean vector for each drug. The portion of the full predicted mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of

Exhibit 8. Elements of Full Predicted Mean Vector

Drug Use Measure and Category of Interest	Predicted Mean
Recency of Use, Past Month[1]	P(past month user | lifetime user)
Recency of Use, Past Year Not Past Month[2]	P(past year but not past month user | lifetime user)
Recency of Use, Past 3 Years Not Past Year[2]	P(past 3 years but not past year user | lifetime user)
12-Month Frequency of Use	P(use on a given day in the year | past year user)[2]*P(past year user | lifetime user)
30-Day Frequency of Use	P(use on a given day in the month | past month user)[2]*P(past month user | lifetime user)
30-Day Binge Drinking Frequency	P(drank 5 or more drinks on a given day in the past month | past month user)[2]*P(past month user | lifetime user)

[1] Note that the final category for recency (lifetime but not past year, or lifetime but not past 3 years) is not needed in the predicted mean vector because the multinomial probabilities add to 1, and this probability is determined by the other probabilities.

[2] Interpreting the proportion of the year used as a probability of use on a given day in the year assumes that the probability of use on each day in the year is equal. This, of course, is not true. However, the violation of this assumption does not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it does allow the predicted mean to be made conditional on what is known.

Exhibit 9. Full Predicted Mean Vector for CAI Sample Drugs

	Drug			
Drug Use Measure and Category of Interest	**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" 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.

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 predicted mean vector used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given in **Appendix H** (see Section H.3). The Mahalanobis distance was then calculated using only the portion of the predicted mean vector that was associated with the given missingness pattern, with elements appropriately adjusted[51]. If no donors were available that 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**.

6.5 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 past NHSDAs; instead, missing values were excluded from subsequent analyses. However, as with the 30-day frequency, missing age at first use values were imputed for the first time in the CAI sample of the 1999 NHSDA and this was continued in the 2000 NHSDA. 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 give 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 were also asked their age at first daily cigarette use.

6.5.1 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 was chosen that was based on a log transformation of the respondent's age at first drug use. UPMNs were formed using the predicted mean from the regression model. Each item nonrespondent's neighborhood was restricted by logical constraints (which cannot be loosened) and likeness constraints (which can

---

[51]See **Appendix C** for a definition of Mahalanobis distance.

be loosened). 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.

6.5.1.1 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 usage hierarchy given in **Exhibit 5**.

6.5.1.2 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, 18 to 25, and 26 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 subset down to lifetime users for each drug. If a valid response was provided for the age at first use measure, 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 (see **Appendix B** for the more general GEM) and included the following categorical covariates: age, race, gender, Census region, an MSA indicator, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

6.5.1.3 Sequential Model Building

After the weight adjustment, the following log transformation was calculated for all lifetime drug users:

$$Y_i = \ln[p_i/1 - p_i], where \qquad p_i = \frac{AgeofFirstUse_i + Uniform(0,1)Number}{(InterviewDate_i - DateofBirth_i)/(365.25)}$$

and where *i* is the drug in question and Y_i is the dependent variable in a weighted linear univariate regression. Variables included in the regression equation[52] were age; age squared; age cubed; State rank (based on the recency variable, see **Section 6.4.1** for details); gender; race/ethnicity; first-order interactions of age, age squared, gender, and race/ethnicity; marital status; educational level; employment status[53]; Census region; an MSA indicator; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; a modified version of the imputed age at first drug use for previously imputed drugs; and modified 12-month and 30-day frequencies for the drug in question. The modified variables for age at first use, 12-month frequency of use (where applicable), and 30-day frequency of use (where applicable) were defined as follows:

new12$_i$ = 0	if respondent did **not** use in the past 12 months
=12-month frequency	if respondent used in the past 12 months
new30$_i$ = 0	if respondent did **not** use in the past month
=30-day frequency	if respondent used in the past month
afu$_i$ =0	if respondent is **not** a lifetime drug user
=age at first use	if respondent is a lifetime drug user

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**.

6.5.1.4 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 defines a "neighborhood" of respondents by requiring that the respondents' predicted age at first use values be 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

---

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

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

available with predicted means within 5 percent of the recipient's predicted mean, the neighborhood was abandoned, and the respondent with the closest predicted was chosen as the donor.

6.5.1.5 Assignment of Imputed Values

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

6.5.1.6 Constraints on Univariate Predictive Mean Neighborhoods

As with all other drug use measures, neighborhoods for age at first use were restricted so that candidate donors and recipients would be 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. In fact, recipients and donors were required to be of the same age, if possible. If a donor could not be found of the same age, 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 could also be considered a likeness constraint, which could be 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, and donors were further required to be the same age as the recipient. Another likeness constraint required that if the item nonrespondent had used the drug in the past year, the donor also had to have used it in the past year. Tobacco users had an additional likeness constraint: If the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years. 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 were more stringent than those for the other drug use measures. It was often necessary, therefore, to loosen the constraints. The order of loosening constraints follows: (1) remove the State rank group; (2) abandon the neighborhood, and choose the donor with the closest predicted mean; (3) remove the requirement that recipients who were users in the past year (or past 3 years for tobacco) had to have donors who used in the past year (or past 3 years for tobacco); (4) loosen the restriction that donors and recipients had to be the same age, and instead require that the donor's age be greater than or equal to the recipient's age and the donor's age at first use be less than or equal to the recipient's age at first use[54]; and (5) loosen the "same-age" restriction even further, so that the donor's age at first use could be less than or equal to the recipient's age. A summary of the above constraints and the number of respondents who fit into each one is listed for each drug in **Appendix G.**

For drugs with no multivariate assignment, there were several logical constraints. Respondents with an age at first use equal to the recipient's current age 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 1st, and the interview date was June 30th. Then the number of days between the interview date and the respondent's birthday would be 90. If the respondent had a 12-month frequency of 100 (either reported or imputed), his or her age at first use could not be his or her current age. In addition, 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 be equal to his or her current age. Finally, 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 be equal to his or her current age. Consider again the example where the recipient respondent's birthday was on March 1st, and the interview was on June 30th, and the number of days between the interview date and the respondent's birthday is 90. If the respondent's 12-month frequency was not a past month user but his or her 12-month frequency was 80, some of those 80 days had to have occurred before his or her birthday, and the respondent's age at first use could not equal his or her current age. Some additional logical constraints were that the donors could not be past year users if the recipient was not a past year user, and, for tobacco, donors could not be users in the past 3 years if the recipient was not a user in the past 3 years. These constraints prevented item nonrespondents from receiving a donated age at first use more recent than the last time they used a substance. 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.

---

[54] 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 equal the current age.

6.5.1.7 Multivariate Assignments

For smokeless tobacco (chewing tobacco and snuff), cocaine (crack), and stimulants (methamphetamines), more than one age at first use variable was associated with a single predicted mean 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.

6.5.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

For reasons discussed in **Section 6.3.7.1**, one model for smokeless tobacco was fit rather than individual models for chewing tobacco and 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. For example, donors for chewing tobacco were logically restricted so that, if the recipient's 12-month chewing tobacco frequency was greater than the number of days since his or her last birthday, donors whose age at first chewing tobacco use was equal to the recipient's age were excluded. The same was true for snuff. As a second example, chewing tobacco donors could not logically be past year chewing tobacco users if recipients were not past year chewing tobacco users. Similar rules applied to snuff (past year and past 3 years) and chewing tobacco (past 3 years). The likeness constraints were also 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 snuff or chewing tobacco in the original data, but the respondent was imputed to be a nonuser of snuff or chewing tobacco in the lifetime imputation, the respondent's age at first snuff use or age at first chewing tobacco use would be adjusted to reflect the situation. Age at first use for smokeless tobacco was obtained by taking the minimum age at first use from snuff and chewing tobacco.

6.5.1.7.2 Cocaine and Crack

Even though cocaine and crack are in distinct modules in the CAI questionnaire, an age at first use model was only fit for 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 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 age at first cocaine use was equal to the recipient's age were excluded. The same was true for crack. As a second example, cocaine donors could not logically be past year cocaine users if recipients were not past year cocaine users. Similar rules applied to past year crack use. The likeness constraints were also 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 be a nonuser of crack in the lifetime imputation, the respondent's age at first crack use would be adjusted to reflect the situation.

Because crack is a type of cocaine, additional logical constraints were required so that donated values would be consistent with preexisting nonmissing values. Specifically, if the crack age at first use was missing and cocaine was not, the donated crack age at first use could not be 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 be 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 be a crack user.

6.5.1.7.3 Stimulants (Methamphetamines and Other Stimulants)

As stated in **Section 6.3.7.3**, the stimulants' module included a subgate question referring to methamphetamines, which is of interest in its own right. One model was fit for stimulants' age at first use, from which a single neighborhood was created for both methamphetamines and stimulants as a whole. The nearest neighbor hot-deck neighborhood was then based on the overall stimulants' predicted age at first use. Missing ages at first use for methamphetamines and/or stimulants as a whole were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and if both methamphetamines and stimulants as a whole 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 stimulants as a whole. Because no 12-month frequency was available for methamphetamines, however, it was not possible to implement any constraints on methamphetamines involving the 12-month frequency.

Because methamphetamines are a type of stimulant, additional logical constraints were required so that donated values would be consistent with preexisting nonmissing values. Specifically, if the age at first use for methamphetamines was missing and overall stimulants was not, the donated methamphetamines' age at first use could not be earlier than the preexisting stimulants' age at first use. Conversely, if the age at first use for stimulants was missing and methamphetamines' age at first use was not, the donated stimulants' age at first use could not be later than preexisting methamphetamines' age at first use. Finally, if the methamphetamines' age at first use was missing but the respondent was a methamphetamines user, the donor had to be a methamphetamines user.

All of the constraints applied specifically to methamphetamines were logical constraints. It is important to note that, for both stimulants and methamphetamines, lifetime usage was considered known (employing the lifetime usage imputation), so that there was no question of use versus nonuse of methamphetamines. If age at first use was missing for methamphetamines in the original data, but the respondent was imputed to be a nonuser of methamphetamines in the lifetime imputation, then the respondent's age at first use of methamphetamines would be adjusted to reflect the situation.

6.5.1.8 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 a nonmissing age at first use value. Using this age at first use (AFU), users were assigned year/month/day of first use values if none was provided. One thing to note is that the day of first use (DFU) was not collected in the questionnaire and was missing for all respondents. Regardless of the number of items missing, all users were assigned a continuous date of first use using either their reported information (for recent initiates) or from a randomly assigned continuous date of first use. The month/day/year were then extracted from this continuous date of first use. The year of first use (YFU), month of first use (MFU), and DFU data contained four patterns of missingness:

1. For less recent initiates: Missing year/month/day of first use (not asked in the CAI instrument: occurs when AFU < current age -1).

2. For recent initiates: Missing month/day of first use (asked in CAI instrument: occurs when AFU = current age or AFU = current age -1).

3. For recent initiates: Missing year/month/day of first use (asked in CAI instrument: occurs when AFU = current age or AFU = current age -1).

4. For recent initiates: Missing day of first use only (asked in CAI instrument: occurs when AFU = current age or AFU = current age -1).

6.5.1.8.1 Missingness Pattern 1

The first type of missingness pattern occurred when the respondent first starting using the drug 2 years or more before his or her current age. This case is analogous to prior year's data where month and year were not asked in the questionnaire. Below is a brief description of the process involved in obtaining a continuous date of first use in such cases. The imputed YFU, MFU, and DFU were extracted from the continuous date defined below.

Continuous date = *Earliest possible date* + [(*Days between earliest and latest date*) * (random #)],

where

Days between earliest and latest = *Latest possible date* - *Earliest possible date*,

Earliest possible date = birth month / birth day / (birth year + age at first use), and

Latest possible date =

minimum [(Interview date - 12 month frequency + 1), (Earliest date + 364 / 365[55])] *if recency = 1*
minimum [(Interview date - 29 - 12-month frequency), (Earliest date + 364 / 365)] *if recency = 2*
minimum [(Interview date - 1 day - 1 year), (Earliest date + 364 / 365)] *if recency = 3*
minimum [(Interview date - 1 day - 3 years), (Earliest date + 364 / 365)] *if recency = 4*

---

[55] This number was changed to 364 if a nonleap year and remained 365 if a leap year.

6.5.1.8.2 Missingness Pattern 2

The second missingness pattern occurred when the respondent recently initiated use (i.e., within 2 years of his or her current age), and the respondent 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 Pattern 1 with the following changes:

- If the *Earliest possible date* < YFU, the *Earliest date* = YFU (using January 1st as the earliest month/day).

- If the *Latest possible date* > YFU, the *Latest date* = YFU (using December 31st as the latest month/day).

6.5.1.8.3 Missingness Pattern 3

Similar to Pattern 2, the third missingness pattern occurred when the respondent recently initiated use (i.e., within 2 years of his or her current age). However, these respondents 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 age at first use. Again, the imputed YFU, MFU, and DFU were derived in the same manner as described in Pattern 1.

6.5.1.8.4 Missingness Pattern 4

In this case, the respondent provided all the information asked by the questionnaire (i.e., both the month and year of first use). However, to obtain a complete date of first use, a day of first use was also needed. Thus, a day of first use was randomly assigned given the respondent's month and year of first use from a uniform distribution in a way that was consistent with both the 12-month frequency/recency and age at first use. Again, the imputed DFU was derived in the same manner as described in Pattern 1 with the following changes:

- If the *Earliest possible date* < reported combination of MFU/YFU, the *Earliest date* = MFU/YFU (using 1st day of the month).

- If the *Latest possible date* > reported combination of MFU/YFU, the *Latest date* = MFU/YFU (using the appropriate last day of the given MFU).

6.5.1.8.5 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. 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 only affected the latest possible date, which was defined as follows for these drugs:

Latest possible date =

minimum [(Interview date - 30-day frequency + 1), (Earliest date + 364/365)] *if recency = 1*
minimum [Interview date - 30), (Earliest date + 364/365)] *if recency = 2*
minimum [(Interview date - 1 day - 1 year), (Earliest date + 364/365)] *if recency = 3*
minimum [(Interview date - 1 day - 3 years), (Earliest date + 364/365)] *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 date was used to produce the smokeless tobacco date of first use. In addition, the combination drugs (i.e., cocaine and crack, stimulants and methamphetamines) had more constraints placed on their assignment of the dates of first use. Because of the complex relationship between these drugs, the cocaine date of first use was made to be consistent with the crack date of first use and vice versa using both cocaine and crack age at first use data, both recency and frequency data, and any given month/year of first use data for either drug (the same was done for stimulants/methamphetamines).

6.5.2 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 cigarette daily 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 one key exception: Whereas the age at first use question was asked of all cigarette users, the age at first daily use question was only asked of 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 answered the "ever-daily-used" question[56] with a "yes," he or she was considered a daily user. 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 had to be imputed.

Thus, the age at first daily use imputation involved two parts. The first part involves missing values in the ever-daily-used question (CG15), which asks the respondent if he or she had ever smoked everyday for at least 30 days, were imputed.[57] The second part involves all missing age at first daily use values for eligible daily users were imputed.

6.5.2.1 Setup for Model Building—Ever-Daily-Used Question (CG15)

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, 18 to 25, and 26 or older), and all subsequent procedures were performed separately within these age groups. To impute for missing values in the ever-daily-used question, it was necessary to define the eligible population: respondents who had an imputation-revised 30-day frequency[58] fewer than 30 days. If a valid response was provided for ever-daily-used question, 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: age, race, gender, Census region, an MSA indicator, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

---

[56] The "ever-daily-used" question is CG15 and was asked of all people who were lifetime but not past month users, or past month users who answered the 30-day frequency (CG07) with a number from 1 to 29. It should have been asked of those with an estimated 30-day frequency (CG07a) that was fewer than 30 (see next footnote).

[57] Besides the traditional source of missing values in CG15 due to answers of "don't know" or "refused," an error in the CAI instrument added another source of missing values. Persons who answered the estimated 30-day frequency with a number smaller than 30 were not given the opportunity to answer CG15 and should have had that opportunity.

[58] The imputation-revised 30-day frequency included responses from the 30-day frequency question (CG07) as well as the estimated 30-day frequency (CG07a).

6.5.2.2 Model Building—Ever-Daily-Used Question (CG15)

After the weights were adjusted, the ever-daily-used question was modeled using weighted logistic regression. Variables included in the initial regression equation were age; age squared; age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of age, age squared, gender, and race/ethnicity; marital status; educational level; employment status[59]; Census region; an MSA indicator; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; a revised 30-day cigarette frequency variable (in the same format as used in the age at first use models, see **Section 6.5.1.3**); and the imputation-revised cigarette age at first use. A summary of the final models can be found in **Appendix F**.

6.5.2.3 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Question (CG15)

From the final model, a predicted mean of the ever-daily-used question 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 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 be within a certain relative distance, delta, of the nonrespondent's predicted probability in order to be included in the neighborhood. Delta was set so that donors were required to have a predicted probability within 5 percent of that of the item nonrespondent.

6.5.2.4 Assignment of Imputed Values—Ever-Daily-Used Question (CG15)

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.

---

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

6.5.2.5 Constraints on Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Question (CG15)

As with all other drug use measures, neighborhoods for the ever-daily-used question were restricted so that candidate donors and recipients would be 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. The likeness constraints were nearly identical to those of age at first use (see **Section 6.5.1.6**). In particular, recipients and donors were required to be of the same age and from States with similar usage levels, if possible. A small delta could also be considered a likeness constraint, which could be 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, and donors were further required to be the same age as the recipient. The recency likeness constraints that were the same as with age at first use for cigarettes follow: (1) if the item nonrespondent had used in the past year, the donor also had to have used in the past year; and (2) if the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years. Two additional likeness constraints were used as logical constraints when they were applied to age at first use for cigarettes: (1) donors could not be past year users if recipients were not past year users; and (2) donors could not be users in the past 3 years if recipients were not users in the past 3 years.

The likeness constraints on the donors were loosened in the following order, until a neighborhood of at least one donor was achieved: (1) remove the State rank group; (2) abandon the neighborhood, and choose the closest predicted mean; (3) remove the requirement that recipients who were users in the past year (or past 3 years for tobacco) had to have donors who used in the past year (or past 3 years for tobacco); (4) loosen the restriction that donors and recipients have to be the same age, so that the donor's age was greater than or equal to the recipient's age; and (5) abandon the "same-age" restriction entirely. To be consistent with the age at first use imputations, the two likeness constraints that were logical constraints in the age at first use imputations were not loosened. A summary of the above constraints, and the number of respondents who fit into each one, is listed for each drug in **Appendix G**.

6.5.2.6 Setup for Model Building—Age at First Daily Cigarette Use

After producing an imputation-revised ever-daily-used variable, the next step was the imputation of age at first daily cigarette use values. The eligible population for age at first daily use incorporates all cases deemed to be daily users for at least 30 days at some point in their lifetime. In other words, eligible respondents either had an imputation-revised 30-day cigarette frequency of 30 days or an imputation-revised ever-daily-used value indicating a period in which the respondent smoked everyday for at least 30 days.[60] The file was broken down into three age categories (12 to 17, 18 to 25, and 26 or older), and all subsequent procedures were performed separately within these age groups. If a valid response was provided for the age at first daily use question, the person was deemed an item respondent. Before modeling, the item respondents' weights were adjusted to match the entire eligible population. These adjusted weights were computed using a response propensity model (see **Appendix B** for the more general GEM) and included the following categorical covariates: age, race, gender, Census region, an MSA indicator, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

After the weights were adjusted, age at first daily cigarette use was modeled using a weighted linear univariate regression with the dependent variable undergoing the same log transformation as the one defined for the age at first use procedure (see **Section 6.5.1.3**). Variables included in the initial regression equation were age; age squared; age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of age, gender, and race/ethnicity; marital status; educational level; employment status[61] ; Census region; MSA; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; modified 30-day cigarette frequency (in the same format as used in the age at first use models); and imputation-revised cigarette age at first use. A summary of the final models can be found in **Appendix F**.

6.5.2.7 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—Age at First Daily Cigarette Use

From the final model, a predicted mean (based on the *Y* variable) was computed for each eligible daily cigarette user. Then a predicted age at first daily use was derived by back-transforming the predicted mean. The imputation-revised age at first daily use assignment was conducted using UPMN imputation. The procedure defines a "neighborhood" of respondents by

[60] Again, incomplete data respondents for the age at first daily use variable included respondents who answered the estimated 30-day frequency as "30" but were not given the opportunity to answer age at first daily use.

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

requiring that the respondent's predicted age at first daily use value be within a certain relative distance, delta, of the nonrespondent's predicted value.

6.5.2.8 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.

6.5.2.9 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 would be 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. The likeness constraints were nearly identical to those of age at first use (see **Section 6.5.1.6**). In particular, recipients and donors were required to be of the same age and from States with similar usage levels, if possible. A small delta could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. Initially, the relative distance for determining age at first daily use imputation neighborhoods (delta) was set so that any potential donor's predicted age at first daily use was within 5 percent of the recipient's predicted age at first daily use, and donors were further required to be the same age as the recipient. The recency likeness constraints were the same as with age at first use for cigarettes: (1) if the item nonrespondent had used in the past year, the donor also had to have used in the past year; and (2) if the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years.

The likeness constraints on the donors were loosened in the following order until a neighborhood of at least one donor was achieved: (1) remove the State rank group; (2) abandon the neighborhood, and choose the donor with the closest predicted mean; (3) remove the requirement that recipients who were users in the past year (or past 3 years for tobacco) had to have donors who used in the past year (or past 3 years for tobacco); (4) loosen the restriction that donors and recipients have to be the same age, and instead require that the donor's age be greater than or equal to the recipient's age and the donor's age at first daily use be less than or equal to the recipient's age at first daily use[62]; and (5) loosen the "same-age" restriction even further, so

[62] 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 equal the current age.

that the donor's age at first daily use could be less than or equal to the recipient's age. A summary of the above constraints, and the number of respondents who fit into each one, is listed for each drug in **Appendix G**.

All the logical constraints applied to cigarettes' age at first use were also applied to age at first daily cigarette use. See **Section 6.5.1.6**, with the words "age at first use" replaced with "age at first daily use." 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.

6.5.2.10 Date of First Daily Cigarette Use Assignments

After the imputation-revised cigarette age at first daily 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. Unlike age at first drug use, the questionnaire did not ask any respondents for their year or month of first daily use of cigarettes. Therefore, the assignment procedure was similar to missing Pattern 1 for age at first drug use (see **Section 6.5.1.8**). Below is a brief description of the process involved in obtaining a continuous date of first daily cigarette use.

Continuous date = *Earliest possible date* + [(*Days between earliest and latest day of first use*) * (random #)]

where

Days between earliest and latest = *Latest possible date* - *Earliest possible date*
Earliest possible date = birth month / birth day/ (birth year + age at first use)
Latest possible date =

minimum [(Interview date - 30-day frequency + 1), (Earliest date + 364/365)] *if recency = 1*
minimum [(Interview date - 30), (Earliest date + 364/365)] *if recency* = 2
minimum [(Interview date - 1 day - 1 year), (Earliest date + 364/365)] *if recency = 3*
minimum [(Interview date - 1 day - 3 years), (Earliest date + 364/365)] *if recency = 4*

From this continuous date of first cigarette daily use, the imputation-revised year/month/day of first daily use was extracted.

7. Health Insurance and Income Imputations

7.1 Introduction

For income and health insurance, several techniques were used to edit and impute missing values. As with some of the demographic imputations in **Chapter 4** and the drug imputations discussed in **Chapter 6**, imputations were accomplished using the predictive mean neighborhood (PMN) technique described in **Appendix C**. However, whereas the editing process for the drug imputations was described in another document (Kroutil, 2002a), the editing procedures implemented on the insurance and income variables are described in this chapter.

7.2 Health Insurance

7.2.1 Edited Insurance Variables

Exhibit 10 summarizes the relationship between a sample of health insurance questionnaire (raw) variables and their edited counterparts. The edited variables have the same values as the questionnaire variables, except that missing values are replaced by standard NHSDA missing value codes. In 2000, two new questions appeared in the NHSDA questionnaire. These questions corresponded to the new variables QHI04 and QHI11, and their respective edited counterparts OGOVTINS and OTHLTINS.

Exhibit 10. Mapping of Raw Health Insurance Variables to Edited Counterparts

Variable	Question	Edited Counterpart
QHI01	Is the respondent currently covered by Medicare?	MEDICARE (1 = yes, 2 = no)
QHI02	Is the respondent currently covered by Medicaid or Medical Assistance?	MEDICAID (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)
QHI04	Is the respondent currently covered by any other government-sponsored program that provides or pays for medical care?	OGOVTINS (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 other type of health insurance or coverage besides the ones I just asked about?	OTHLTINS (1 = yes, 2 = no)

Three health insurance indicators were created from these six variables. Two of them, INSUR and INSUR2, indicate whether the respondent has any health insurance; the third, PINSUR, indicates whether the respondent has any private health insurance. INSUR2 was coded as "yes" if any one of the six variables listed in **Exhibit 10** were coded as "yes," and it was coded as "no" if all six variables were coded as "no." The other overall insurance indicator, INSUR, was created to maintain consistency with 1999. Because the questions associated with OGOVTINS and OTHLTINS did not exist on the 1999 questionnaire, these two variables were excluded from the determination of INSUR, which was coded as "yes" if any of the other four variables listed in **Exhibit 10** were coded as "yes," and it was coded as "no" if all four variables were coded as "no." The variable PINSUR used only PRVHLTIN. Missing data for the edited variable PRVHLTIN were coded using the standard NHSDA missing data codes for "don't know," refused, and blank, whereas missing data for PINSUR were all coded as "98." 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 NHSDAs in which other variables also contributed to the indicator of coverage by private health insurance. All respondents with private health insurance were considered to have health insurance; therefore, respondents with private health insurance are a subset of the respondents who had health insurance.

7.2.2 Imputed Health Insurance Variables

7.2.2.1 Order of Modeling Health Insurance Variables

A multivariate predictive mean neighborhood imputation method for private health insurance and overall health insurance was implemented. However, respondents who answered "yes" to the private health insurance question were logically also covered by overall health insurance. Therefore, it was not possible to use INSUR as a covariate in the PINSUR model, or vice versa. Consequently, the models for the two variables could be run simultaneously.

7.2.2.2 Setup for Model Building

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 both health insurance variables were conducted separately within the four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older.

In 2000, an interview respondent was considered an item respondent for health insurance only if he or she was a respondent to both overall health insurance and private health insurance. To meet this criterion, the interview respondent must have answered QHI06 and either must have answered "no" to all six questions or "yes" to at least one of the six questions. This ensured that the interview respondent's status with respect to both overall health insurance and private health insurance is 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 INSUR was already known to be a "yes."

In the 1999 CAI sample, respondents and nonrespondents were determined separately for the two health insurance variables. Also in 1999, response propensity adjustments to the weights were implemented independently for the two variables within each age group. Due to making minor modifications to the program regarding the definitions of respondents and nonrespondents, only one response propensity program was needed in 2000 for both health insurance variables within each age group.

These programs contain the item response propensity model, which is a special case of the generalized exponential model (GEM). Greater details of the GEM model are presented 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.

7.2.2.3 Sequential Model Building

The probability that the respondent had health insurance 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. For the models, the parameters were estimated using logistic regression. Each response propensity model appraised predictors, which included continuous age, race/ethnicity, age squared, gender, population density, percentage of housing in segment that is owner-occupied, percentage concentration of Hispanics in the segment, percentage concentration of non-Hispanic blacks in the segment, and household size. There were also predictors that consisted of one-way interactions of age with race/ethnicity, age with gender, race/ethnicity with gender, age squared with race/ethnicity, and age squared with gender. For the three older age groups (i.e., 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older), the additional predictors of marital status, educational level, and employment status were also considered in each model.

7.2.2.4 Computation of Predictive Means

Using the parameter estimates from models for overall and private health insurance, predicted probabilities of use were computed for both item respondents and nonrespondents. Because neither variable could be used as a covariate in the model for the other variable, no provisional values were required.

7.2.2.5 Multivariate Imputation of Health Insurance and Private Health Insurance

The final imputed values for health insurance and private health insurance were obtained using neighborhoods built upon a vector of predictive means. For both overall and private health insurance, the imputation method used was the multivariate predictive mean neighborhood (MPMN) model-based 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 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older.

A respondent was eligible to be a donor for a given item nonrespondent if he or she had complete data across both health insurance variables and was within the same age group. Logical constraints were placed on individuals who were missing one indicator but not the other. Respondents who were missing the overall indicator, but did not have private health insurance, required donors who also did not have private health insurance.[63] Likewise, respondents who indicated that they had health insurance but were missing the private health insurance indicator required donors who had some health insurance.[64] As a likeness constraint, the set of potential donors was then further restricted to be 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 each element of the multivariate predicted mean vector "close to" (i.e., within the delta distance) the recipient's elements of the predicted mean vector. Because the imputation was multivariate, the set of deltas was also multivariate,

---

[63] Technically, this is not a logical constraint because there is no restriction on whether the respondent does or does not have health insurance. However, because all respondents with private health insurance have health insurance, and the recipient does not have private health insurance, the distribution would be skewed in favor of a "yes" indicator if these respondents were allowed to be donors.

[64] As with the previous footnote, this technically is not a logical constraint. However, because all respondents who do not have health insurance also do not have private health insurance, and the recipient has health insurance, the distribution would be skewed in favor of a "no" indicator if these respondents were allowed to be donors.

where a different delta corresponded to each element of the predicted mean vector. Likeness constraints were loosened in the order given 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 given 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 and for private health insurance. 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 one of the two health insurance variables was not missing, the predictive mean vector used to determine the neighborhood was limited to the predictive mean associated with the missing variable. The portions of the full predictive mean vector used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given 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. If no donors were available that 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**.

7.3 Income

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 referred to in the previous questions). The correspondence between these questionnaire items and the edited variables is given in **Exhibit 11**. The second phase of the imputation of income was known as the "specific category phase" and consisted of imputing more specific income categories for the respondent and the respondent's family in the household.

Exhibit 11. Mapping of Questionnaire Income Variables to Edited Counterparts

Source of Income/Binary Total Income Questions				
Variable Description	**Raw Questions**	**Edited Personal Income**	**Edited Other Family Income**	**Edited Total Family Income**
Social Security	QI01, QI02	PSOC	OFMSOC	FAMSOC
Supplemental Security	QI03, QI04A, QI04B	PSSI	OFMSSI	FAMSSI
Wages	QI05, QI06A, QI06B	PWAG	OFMWAG	FAMWAG
Food Stamps	QI07A, QI07B	-----*	-----*	FSTAMP
Welfare Payments	QI08, QI09A, QI09B	PPMT	OFMPMT	FAMPMT
Other Welfare Services	QI10, QI11A, QI11B	PSVC	OFMSVC	FAMSVC
Months on Welfare	QI12A, QI12B	-----*	-----*	WELMOS
Investment Income	QI13, QI14A, QI14B	PINT	OFMINT	FAMINT
Child Support	QI15, QI16A, QI16B	PCHD	OFMCHD	FAMCHD
Other Income	QI17, QI18A, QI18B	POTH	OFMOTH	FAMOTH
Total Income	QI20, QI22	PINC1	FINC1[1]	FAMINC1
Total Income Specific Categories	QI21A, QI21B, QI23A, QI23B	PINC2	FINC2[1]	FAMINC2

* Edited variables are not generated.

[1] These variables apply to all family members in the household. They include the respondent, unlike the OFM*xxx* variables.

7.3.1 Edited Income Variables: Binary Variable Phase

7.3.1.1 Source of Income Variables

Most of the variables measuring the source of income consisted of two parts, which were personal source of income and other-family-member source of income. The first questions asked whether the respondent received income from a particular source. If the response was "yes" or if the respondent did not have other family members in the household, the other-family-member question should have been skipped.[65] From these two parts, three edited income

[65] The CAI logic routed the respondent to the other-family-member question only if family relationship codes were present in the household roster. There were instances, however, when family relationship codes were in the household roster, but were set to missing in the roster edits (see **Chapter 8**) due to logical inconsistencies. It is

source variables were created. These edited variables were personal source of income, other-family-member source of income, and total family source of income. Among the source of income variables, exceptions to this paired question format included questions regarding food stamps and the number of months on welfare. For these questions, only one question was asked, which applied to the entire family in the respondent's household.

Every respondent was eligible to answer the personal source of income questions. Hence, the raw and edited personal source of income variables are equivalent. Yet the other-family-member income questions required more editing. As stated previously, if the respondent answered "yes" to the personal question or did not have any family members in the household, the other-family-member question should have been skipped and was coded as a legitimate skip.[66] If the respondent was not skipped out of the other-family-member question, he or she was asked either the A or B version of the question depending on the answers to previous personal income questions. Editing was conducted to merge these A and B questions into one other-family-member source of income variable.

Edited variables were not generated for some of the personal income and some of the other family income. For instance, food stamps information was collected using one question (QI07A/B) that applied to the respondent's entire family. Also, the question concerning months on welfare (QI12A/B) was only asked for respondents who answered "yes" to either the welfare payments (personal: QI08, or other family: QI09A/B) or other welfare services (personal: QI10, or other family: QI11A/B) source of income questions.

7.3.1.2 Personal and Family Total Income Variables

In addition to the source of income variables, the binary variable phase also included a pair of binary variables regarding whether the respondent's personal total income or the respondent's family's total income was $20,000 or more. For this pair of questions (QI21 A/B; QI23 A/B), the second question in the pair applied to the entire family. As with the source of income variables, the raw and edited personal total income variables were equivalent. The second question in the pair asked about total family income, but was skipped if the respondent

---

possible that the family skip variable (IRFAMSKP) would have then been imputed to indicate that no other family members were present in the household, even though the other-family-member question had data in it.

[66] When the family skip variable IRFAMSKP indicated no other family members in the household, but the respondent was routed to the other-family-member question because of his or her roster information, the legitimate skip that would be coded in the other-family-member variable would overwrite real data, rather than an NHSDA blank data code. However, such cases rarely occurred.

had no other family members in the household. The edited variable FINC1 was created by assigning legitimate skips in those cases. A third binary total family income variable FAMINC1 was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household. Finally, if the total personal income response indicated an income of $20,000 or more, but the total family income response was less than $20,000, the values for all three variables were set to missing and later imputed.

7.3.2 Imputed Income Variables: Binary Variable Phase

7.3.2.1 Order of Modeling Income Variables

After editing the income variables, the next step in the imputation of income variables was to determine the order in which the variables would be modeled. Greater details of the hierarchy in which the income variables are modeled are provided in **Appendix C**. 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. The resulting values were temporary at this stage. This was due to the fact that 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 **Exhibit 12**.

7.3.2.2 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 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. For an individual to be considered an 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 social 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 vs. $20,000 or more). Response

Exhibit 12. Order of Imputation of Income Variables in Binary Variable Phase and Response Variables Used in Models

Income	Edited Family Variables
Social Security	FAMSOC
Supplemental Social Security	FAMSSI
Welfare Payments	FAMPMT
Other Welfare Services	FAMSVC
Investment Income	FAMINT
Child Support Payments	FAMCHD
Wages	FAMWAG
Other Income	FAMOTH
Food Stamps	FSTAMP
Months on Welfare	WELMOS
Total Family Income[1]	FINC1

[1] Total family income uses all of the predictors mentioned above except months on welfare.

propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. Since item respondents were defined across all the income variables in the binary variable phase, this adjustment was only computed once per age group and then used in the modeling of income indicators. The item response propensity model is a special case of GEM, which is described in greater detail in **Appendix B**. The model variables, which predicted the probability of item nonresponse, were the same as those included in the main model, which is discussed in the next section.

7.3.2.3 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. 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 **Exhibit 12**. The covariates in each response propensity model were continuous age, age squared, gender, race/ethnicity, provisional income indicators earlier in the sequence, region, population density, percent Hispanic population, percent non-Hispanic black population, 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 age with race/ethnicity, age with gender, race/ethnicity with gender, age squared with race/ethnicity, and 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 whose income was greater than or equal to \$20,000.

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,[67] where $Y = \text{logit}(p)$ 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**.

7.3.2.4 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, predictive means needed to be calculated 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 UPMN imputation method described in **Appendix C**.

7.3.2.5 Assignment of Provisional Imputed Values

Separate assignments of provisional values were performed within each of the four age groups for all income variables. The final income imputations were multivariate across all the variables in the binary variable phase. These variables consisted of source of income, months on welfare, and the total income variables. The multivariate imputation process is further described in **Section 7.3.2.8**.

---

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

7.3.2.6 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 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. As a logical constraint, 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 be within the same age group (12 to 17, 18 to 25, 26 to 64, 65 or older). Models were built separately within these four groups, so this likeness constraint was never loosened. A small delta could also be considered a likeness constraint, which could be loosened by enlarging delta, or abandoning the neighborhood altogether and taking the donor with the closest predicted mean. This was the only likeness constraint that could be loosened with the binary income provisional imputations.

7.3.2.7 Multivariate Assignments

The predictive means were calculated with edited family income variables (see **Exhibit 12**) 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. These two variables were food stamps and months on welfare. The variables associated with each of the models are given in **Exhibit 13.**

7.3.2.8 Multivariate Imputation

Sections 7.3.2.1 through **7.3.2.7** summarize the specifics of separating the set of income variables (in the 2000 NHSDA) 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

Exhibit 13. Imputation-Revised Personal and Family Income Variables

Income Model	Variables
Social Security	IRPSOC, IROFMSOC, IRFAMSOC
Supplemental Social Security	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

measures were obtained using neighborhoods built on a vector of predictive means using the MPMN technique described in **Appendix C**. Consistent with the univariate imputations, the multivariate assignments were done separately within four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older.

The source-of-income variables, a single months-on-welfare variable, and the binary total income variables are outlined in **Exhibit 11**. The collective distance between these variables' conditional predictive means for a given incomplete data respondent and the complete data respondents was determined using a Mahalanobis distance within each age group. As with other applications of MPMN, the predictive mean vector used in the Mahalanobis distance calculation only included 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 only missing a response for the other-family welfare payments question, the donor's other-family welfare payments response was given to the recipient, as well as the family welfare payments variable IRFAMPMT.

Candidate donors were restricted according to logical constraints, which cannot be loosened. As with the univariate provisional imputations, donors and recipients are required, as a logical constraint, to have 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 were also imposed on the multivariate neighborhood for the binary income variables. The donors were usually restricted to have an age the same as the recipient, or if that constraint was too restrictive, an age within 5 years of the recipient was used. Of the variables outlined in **Exhibit 11**, there was a high degree of association between respondents who received welfare, 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. If insufficient donors were present, the constraints were loosened in the following order: (1) abandon the neighborhood, and choose the donor with the closest predicted mean; (2) remove the requirement that donor and recipient be of the same age, but require them to be within 5 years of each other; (3) remove the requirement that the donor and recipient have ages within 5 years of each other; then (4) remove the constraint that incorporated the association between the welfare, food stamps, and income payment questions. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix G**.

7.3.3 Edited Income Variables: Specific Category Phase

As part of the second phase of the income questions, respondents were asked to identify, both for themselves and for their families, specific categories of income, within the two general categories previously selected. The first general income category consisted of less than $20,000, while the second one consisted of $20,000 or more. In particular, for respondents who answered the binary total income question as less than $20,000, they were asked to enter a specific category of income within increments of $1,000 up to $20,000. These increments consisted of ranges between $0 and $999, $1,000 and $1,999, and others up to a range of $19,000 and $19,999. Conversely, respondents who answered the binary total income question as $20,000 or more were asked to enter a specific category of income within increments of $5,000 up to $50,000. These increments consisted of ranges between $20,000 and $24,999, $25,000 and $29,999, and others up to a range of $45,000 to $49,999. If the respondent's income was greater than $50,000, he or she had a choice of selecting between $50,000 and $74,999 or more than $75,000.

As with the binary total income questions, the specific category questions were asked in a pair, the first for the individual respondent and the second for the entire family. Like other variables that followed this pair pattern, the raw and edited personal total income variables were equivalent. The second question was skipped if the respondent had no other family members in the household.[68] The edited variable was created by assigning legitimate skips in those cases. A third specific category family total income variable was created, which would be equal to the response to the second question in the pair if other family members were present in the household. Conversely, if no other family members were present, this family total income variable was equal to the response to the first question in the pair that related to the individual respondent. Finally, if the binary total income responses were set to bad data, the specific category responses were also set to bad data.

7.3.4 Imputed Income Variables: Specific Category Phase

7.3.4.1 Hierarchy of Income Variables

Three income variables resulted from editing the questions in the income-specific category phase (see **Exhibit 11**). These three variables were all considered simultaneously using a failure time model, which is described in greater detail in **Section 7.3.4.3**. Because only one model was fit, no hierarchy was required.

7.3.4.2 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 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. For an individual to be considered an item respondent for income variables in the specific 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 in order to make the item respondent weights representative of the entire sample, and the appropriately adjusted weights were used in the models. The item response propensity model is a special case of the GEM, which is described in greater detail in

---

[68] If no family relationship codes were present in the household roster, the respondent was automatically skipped out of the question about family income. There were instances, however, when family relationship codes in the household roster did not make any sense. The CAI logic would still route the respondent to the family income question. However, in the CAI roster edits, the family relationship codes would be set to bad data (see **Chapter 8**). It is possible that the family skip variable (IRFAMSKP) would be then imputed to indicate that no other family members were present in the household. Hence, the legitimate skip coded in the family income variable would overwrite real data rather than a NHSDA blank data code. However, such cases rarely occurred.

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.

7.3.4.3 Sequential Model Building

The specific categories of income were modeled using the LIFEREG procedure in SAS.[69] This procedure was used for regression modeling of continuous non-negative 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 y, which in this case represents income, is

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

where y is a vector of observed responses, $\boldsymbol{X}$ is the matrix of covariates, β is the parameter vector, and ε 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 NHSDA income intervals. The contribution of an individual with covariates in the matrix $\boldsymbol{X}$ to the overall likelihood is just 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 $F(u|\boldsymbol{X},\beta,\sigma) - F(l|\boldsymbol{X},\beta,\sigma)$, where F is a cumulative distribution function. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.[70]

LIFEREG allowed several choices for the functional form of the parametric model that correspond 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-

---

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

[70] Details about the model specifications for LIFEREG models are given in the *SAS/STAT User's Guide, Version 8* (SAS Institute, 1999, pp. 1761-1796).

parameter special cases when using the likelihood ratio tests as a criteria for comparison, it was decided to use a two-parameter model.

Many of the covariates considered in the model for the specific category phase included the same covariates used in the binary variable phase. These covariates included continuous age, age squared, gender, race/ethnicity, region, population density, percent Hispanic population, percent non-Hispanic black population, percent 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 specific category group were defined in terms of the proportion of a given State's residents whose income was greater than or equal to $20,000. For both phases, there were also predictors that consisted of one-way interactions of age with race/ethnicity, age with gender, race/ethnicity with gender, age squared with race/ethnicity, and 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 specific category phase. Also, all imputation-revised income indicators considered in the binary variable phase were used as covariates for the specific category phase.

7.3.4.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods

As described in the previous section, the failure time model contained the term $\boldsymbol{X}\beta$, which was the predictive mean value. This value was a monotonic function of the conditional mean of the modeled income distribution at a given individual set of values of the regressor variables. Specifically, $\boldsymbol{X}\beta$ 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 **Appendix D**.

7.3.4.5 Assignment of Imputed Values

Separate assignments of imputed values were performed within each of the four age groups for all specific category income variables. Only missing values were replaced by imputed values using the same donor for all three variables. The multivariate imputation process is further described in **Section 7.3.4.7**.

7.3.4.6 Constraints on Univariate Predictive Mean Neighborhoods

Donors and recipients were required to have the same values for both the binary income variable and the indicator of whether other family members were in the household (IRFAMSKP). In addition, if either of the personal income or family income specific 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 specific 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**.

7.3.4.7 Multivariate Assignments

The predictive means were calculated using the edited (specific category) family income variables (see **Exhibit 12**) 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 missing values for 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. The imputation-revised variable for the personal income variable is called IRPINC2, the family income variable with legitimate skips is called IRFINC2, and the family income variable without legitimate skips is called IRFAMIN2.

8. Household Composition (Roster) Editing and Imputations

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. As with the drug imputations discussed in a previous chapter (**Chapter 6**), imputations were accomplished using the predictive mean neighborhood (PMN) technique described in **Appendix C**. However, whereas the editing process for the drug imputations are described elsewhere (see Kroutil, 2002a), the editing procedures implemented on the household roster, the procedures to create respondent-level detailed roster variables, and the procedures to create the roster-derived household composition 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 questionnaire household roster (QD54) asked the respondent (interviewer administered) 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 were 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 filled out a grid with the number of rows corresponding to the value entered in QD54. An example of such a grid when QD54 = 4 is given in **Exhibit 14**. In this example, the roster of the wife/mother is given, and an indicator says that the other pair member selected was the son. The relationship codes are given in **Exhibit 15**. Also given in **Exhibit 15** are details corresponding to certain relationship codes.

Exhibit 14. Household Composition (Roster) Grid Example, QD54 = 4

Person #	Relationship to Respondent	Age in Years	Other Member Selected
1	Self	44	0 (No [Impossible])
2	Husband	42	0 (No)
3	Son	16	1 (Yes)
4	Boarder/Roomer	16	0 (No)

Exhibit 15. 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	Living Together as Though Married	
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 Preliminary Roster Edits

To facilitate processing of the roster variables, a "roster-level" file was created in which the number of records per respondent is given by the household size in QD54. If the respondent broke off 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.

8.2.3 Roster Edits Involving the Self

If only one roster member was identified as "self," where the age of the roster member was within 1 year of the questionnaire-edited age[71] (AGE, defined in **Chapter 4**), and the gender for self matched IRSEX (also defined in **Chapter 4**), the roster age was set to AGE, and no

[71] A 1-year difference was allowed because a respondent's age might have changed during the interview.

further action was required for the self record. However, there were three ways in which an interviewer could enter incorrect information for the self in the household roster: (1) no self in roster, (2) multiple selves in roster, or (3) the roster age for self differed from AGE by more than 1 year, or the gender for self in the roster did not match IRSEX (also defined in **Chapter 4**). Each of these "self" edits is discussed below.

8.2.3.1 Edits for No Self in Roster

If no self was identified in the roster, an attempt was made to identify a self among the roster members corresponding to the respondent in question. 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, and had a relationship code that was impossible. Only one roster member had a relationship code changed to self, so among all the roster-level records corresponding to the respondent, the self code was assigned to the roster member in the following priority order: (1) the roster member was the respondent's biological, adoptive, or foster parent, but was within 1 year in age of the respondent and was the same gender as the respondent; (2) the roster member was younger than 15 years old and was within 1 year in age of the respondent, but was the respondent's parent (the roster member and respondent also had the same gender); (3) the roster member was the respondent's biological, adoptive, or foster child, but was within 1 year in age of the respondent and was the same gender as the respondent; (4) the roster member was the respondent's child, but the respondent was younger than 15, which was within 1 year in age of the roster member (the roster member and respondent also had the same gender); (5) the roster member was the respondent's spouse (not live-in partner), grandchild, or grandparent, but was within 1 year in age of the respondent and was the same gender as the respondent; (6) the roster member's relationship, age, and gender data were missing. If the roster member's relationship code, age, and gender data were missing, the relationship code was set to self, the roster age set to AGE, and the roster gender set to IRSEX. If no roster member met the above criteria, it was assumed that the respondent did not consider himself or herself when counting the number of people in his or her household. The value of QD54 was assumed to be wrong (one fewer than necessary), and a record was added with a relationship code of self, a roster age equal to AGE, and a roster gender equal to IRSEX.

8.2.3.2 Edits for Multiple Selves in Roster

If multiple selves were identified in the roster, an attempt was made to identify the correct self among all roster members with a "self" relationship code. If one or more of the roster members with the self code had a roster age that matched the edited questionnaire age (AGE) exactly, and roster gender matched IRSEX, the first among these roster members was selected as the true self. If no exact match was available, but one or more of the roster members with the self code had a roster age that differed from AGE by a year, with an exact match on IRSEX, the first among these roster members was selected as the true self. Finally, if none of the roster members with the self relationship code had an age-gender approximate match (age within 1 year) with AGE and IRSEX, QD54 was assumed to be wrong (one fewer than necessary), and a record was added with the relationship code of self, a roster age equal to AGE, and a roster gender equal to IRSEX.

8.2.3.3 Edits for Cases When the Assigned Self Did Not Have Appropriate Age or Gender

Although the interviewer might have identified a single roster member as the self, it was possible that the identification was incorrect and that the self may actually have corresponded to a different roster member. Perhaps the interviewer may have applied the wrong relationship codes to the roster members using a household member other than the respondent as the reference point. Using the example given in **Exhibit 14**, if the respondent's son was used as the reference point, the relationship for the respondent became "mother" instead of "self" and the husband became "father." Under these circumstances, the self code was set to missing, and the respondent's roster entries became a no-self household. The procedures for finding the roster member who was the self was then equivalent to the no-self case outlined in **Section 8.2.3.1**.

8.2.4 Roster Edits for Other Household Members

Relationship codes were edited if the relationship of the roster member was impossible based on age and gender, and a self code was not assigned. If the household roster originally did not have a self, candidates for the self were selected among cases where the given relationship code was impossible, as discussed in **Section 8.2.3.1**. If more than one roster member had impossible relationship codes according to the criteria given in that section, then the roster members not assigned a self code were given a bad data code. Otherwise, edits of roster ages, genders, and/or relationship codes either changed the value to another value, or changed the 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.

8.2.4.1 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, where the age, gender, and/or relationship code given was/were either missing or internally inconsistent, and replaced by (an) internally consistent value(s). 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 not uncommon for a double-digit age to be entered as a single-digit age ("5" instead of "55"), or vice versa ("55" instead of "5"). If the relationship code was not nonsensical (e.g., "other relative"), this type of error was difficult to detect. Even if such errors were accompanied by a nonsensical relationship code, this does not generate a problem with the CAI program, and it does not flag such relationships as a 4-year old parent of a 17 year old. In this example, it would be difficult to say whether the error was due to the age or the relationship code. However, if two pair members were selected in a household, these errors can be detected and corrected by observing 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 a *xx*-year-old and no *x*-year-old, where "x" denoted a single-digit number, it was highly probable that an error such as this had occurred. By looking at the number of children under 12 in each roster and comparing it with the screener roster, it became readily apparent whether and how a correction should be made. In this instance, the offending age was replaced by the value given by the pair member with the roster agreeing with the screener.

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.

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.

4. The relationship code for grandchild (9) and grandparent (11) were commonly confused. 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, the relationship code was changed to grandparent.

8.2.4.2 Edits to Relationship Codes: Changes to Missing Codes

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent, and no internally consistent value could be used to replace it. These edits were performed after the edits in **Section 8.2.4.1**. The relationship code in this instance was set to a bad data code.

1. More than one roster member was listed as being the respondent's spouse or as living together with the respondent as though married. For all roster members with such relationship codes, the relationship codes were set to missing.

2. The roster member was the respondent's biological, adoptive, or foster parent, but was younger than the respondent.

3. The roster member was the respondent's step-parent, but was younger than 18, and was at least 20 years younger than the respondent.

4. The roster member was the respondent's biological parent, but was fewer than 12 years older than the respondent.

5. The roster member was the respondent's biological mother, but was more than 60 years older than the respondent.

6. The roster member was the respondent's parent, but was younger than or the same age as the respondent and was under 18 years of age.

7. The roster member was the respondent's biological, adoptive, or foster child, but was older than the respondent.

8. The roster member was the respondent's stepchild, but was at least 20 years older than the respondent, and the respondent was under 18.

9. The roster member was the respondent's biological child, but was fewer than 12 years younger than the respondent.

10. A respondent had a biological sibling older than a biological parent. If this occurred, the relationship codes of both the "sibling" and the "parent" were set to missing.

11. The roster member was the respondent's parent-in-law or child-in-law, but the either the roster member or the respondent was under 15 years old.

12. The roster member was the respondent's child-in-law, but was at least 10 years older than the respondent.

13. The roster member was the respondent's parent-in-law, but was at least 10 years younger than the respondent.

14. The roster member was the respondent's child-in-law, but the child-in-law was under 15 years old. (If the respondent was older than 25, the code was set to child rather than to missing.)

15. The respondent had two children-in-law, but no children in the household. The in-law codes were set to missing.

16. The roster member was the respondent's grandchild, but the roster member was older than the respondent or the respondent was 25 years old or younger.

17. The roster member was the respondent's grandchild, but the respondent's parents lived in the household, the respondent had no children in the household, the respondent was less than 24 years older than the roster member.

18. The roster member was the respondent's sibling (biological, adoptive, or foster), but the roster member's age was within 4 years of the age of the oldest parent.

19. The roster member was the respondent's step-sibling, but the roster member's age was within 4 years of the age of the parent, of which there was only one.

20. The roster member was the respondent's grandparent or grandchild, but the age difference between the respondent and the roster member was under 20 years.

In addition, 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.

8.2.4.3 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Nonsensical Child Code)

In the previous section (**Section 8.2.4.2**), nonsensical relationship codes were set to bad data. Often this occurs 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 can be used to determine the appropriate relationship code for that individual. Scenarios where the nonsensical code was "child" are listed below.

1. The interviewer might 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 would be nonsensical if the age made it impossible for the roster member to be the respondent's child. (See #9 in **Section 8.2.4.2**. In fact, more than one "child" could be listed after the respondent's parent, each of which could be listed as nonsensical.) 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, the relationship code was changed to sibling. Similarly, if the respondent was unmarried and not living with a partner, and the roster member was not 12 or more years younger than the respondent, the relationship code was changed to sibling.

2. Both sides in a selected pair were respondents under 18, both sides identified parents in the household, and one side had a nonsensical child code. When the number of nonsensical 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. The relationship code was changed to sibling.

3. One side of a selected pair listed as the respondent's child, a roster member that was not more than 12 years younger than the respondent, and the respondent was 25 or younger. The relationship code was listed as child, and the previous roster member was listed as grandparent. The "child" was in reference to the respondent's grandparent and was 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 non-immediate family codes (7, 12, 13, or 14) on either side of a selected pair, then no uncles/aunts live in the household. Otherwise, one could not be sure, so the relationship code was set to missing.

8.2.4.4 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Nonsensical Spouse Code)

The interviewer also could have used a wrong reference person with spouse codes. The most common manifestation of this type occurred when a selected child had a parent with a spouse (the other parent) or live-in romantic partner ("living together as though married"). Rather than identifying this individual as a "parent" or "other non-relative," the interviewer identified the roster member as a spouse or live-in romantic partner of the child, even though they intended for the point of reference to be the child's parent rather than the child. This manifestation of the invalid spouse code, along with others, is given below.

1. Both sides in a selected pair identify a spouse, one respondent was much older than the other, and the younger respondent had an unusually large age difference between the respondent and the "spouse." If the younger respondent indicated a parent and the older respondent indicated neither parents nor parents-in-law, the older respondent was either the younger respondent's parent or the parent’s spouse. The misidentified spouse of the younger respondent was either the respondent's parent or the parent's spouse/live-in partner.

2. Both sides in a selected pair identified a spouse, both were under 21, and both had unusually large age differences between the respondents and their "spouses." If both respondents indicated a parent in the household, the respondents were siblings, and on each side the misidentified spouse was a spouse of the respondent's parent. As above, the misidentified spouse was either the respondent's parent or the parent's spouse/live-in partner.

3. Both sides in a selected pair identified a spouse, one respondent was much older than the other, and either the younger respondent was under 21 or had a spouse older than his/her parent. If the previous roster member in the younger respondent's roster was the parent, then the "spouse" code was in reference to the respondent's parent.

4. A spouse (not live-in partner) was identified even though either (1) the respondent was under 15; (2) the spouse was under 15; or (3) the respondent was under 18, but says he/she was "never married" in the core part of the questionnaire. If the young respondent listed one parent, but the other pair member listed two parents, the relationship code was in reference to the parent. If the respondent listed one fewer sibling than the other respondent, the spouse code was a typographical error, meant to be a sibling (4).

5. Both sides identified the same household member as spouse. If the previous roster member on one of the sides was a sibling, the spouse was the sibling's spouse. The relationship code was changed to "other relative."

6. Only one respondent was selected in the household, either the respondent was under 21 or had a spouse (not live-in partner) older than his/her parent, and the age difference between the respondent and the spouse was large. If the previous roster entry was the respondent's parent, the spouse was in reference to the parent. The misidentified spouse of the respondent was the respondent's parent.

7. Only one respondent was selected in the household, either the respondent was under 15 or had a spouse or live-in partner older than his/her parent, and the age difference between the respondent and the live-in partner was greater than 12 years. If the previous roster entry was the respondent's parent, the live-in partner was in reference to the parent. The misidentified live-in partner of the respondent was set to "parent."

8. In all other cases where the respondent was under 15, and identified a spouse, the relationship code was set to bad data.

In most cases where the misidentified spouse was the respondent's parent's spouse or live-in partner, the code was changed to parent. The exception occurred when (1) the respondent with the misidentified code was under 18; (2) the live-in partner of this respondent's parent was the other respondent selected in a pair; and (3) the live-in partner did not indicate that the other pair member selected was his/her child in the parenting experiences question, FIPE3.

8.2.4.5 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Nonsensical 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, the "sibling" relationship code was suspicious. If the previous roster entry was either a parent 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, the sibling relationship codes were referencing the respondent's children's relationships to each other. The relationship codes were therefore changed to "child." Rosters with age differences between 20 and 25 years were individually checked to make sure this change was reasonable.

8.2.4.6 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Nonsensical 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 is 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, the grandchild was in fact the respondent's sibling. However, if this could not be established, the roster member would be the respondent's sibling or the respondent's cousin, so the code was set to bad data.

8.2.4.7 Edits to Relationship Codes: Changes to Different Values (Invalid Reference Person: Nonsensical 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/herself as the parent-in-law of the roster member.

1. One side listed as the respondent's child-in-law, a roster member who was not more than 12 years younger than the respondent, and the respondent was 25 or younger. The relationship code was listed as child-in-law, and the previous roster member was listed as grandparent. The "child-in-law" was in reference to the respondent's grandparent and was 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 non-immediate family codes on either side of a selected pair, no uncles/aunts live in the household. Otherwise, one could not be sure, so the relationship code was set to missing.

2. A respondent's "parent-in-law" was at least 10 years younger than the respondent, and the respondent had a child. It would be logical to assume that the respondent's parent-in-law was in fact the child's spouse, and the code for parent-in-law was changed to child-in-law.

3. A respondent's "child-in-law" was at least 10 years older than the respondent, and the respondent had a spouse. It would be logical to assume that the respondent's child-in-law was in fact the spouse's parent, and the code for child-in-law was changed to parent-in-law.

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 sex), ROSRLT1-ROSRLT25 (relationship to respondent), and ROSMSL1-ROSMSL25 (0/1 indicator: other member selected).

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 **Exhibit 16**. 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, roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could be imputed. Details on the imputation of the household roster-derived variables are given in **Section 8.5**.

The respondent's household size was assumed to equal the total number of rostered people in the household, TOTPEOP, as shown in **Exhibit 16**. The value of TOTPEOP was expected to equal to QD54 in most cases. However, in some cases the assigned self did not match, even approximately, the respondent's age or gender, or no self was assigned and no other roster members matched 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 one greater than QD54. In some cases, the respondent did not enter a value for QD54, so that TOTPEOP and all the roster-derived variables were missing.

KID17 (number of children in the household under 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

Exhibit 16. 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 (not on public use file)	FAMSKIP
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 less 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)	IFATHER

[1] The IMOTHER and IFATHER indicators are not 0/1 indicators because levels are provided for "unknown" and "18 or over."

members had missing ages, the values of KID17 and HH65 would be missing, regardless of whether some of the roster members were eligible to be part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in **Section 8.5.6**. However, if the roster member was missing a relationship code, but not an age, that roster member was still eligible to be counted in these variables.

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, the value of FAMSKIP would be 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, the value of FAMSKIP would not be missing even if other roster members had missing relationship codes.

Nine 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 **Exhibit 16**. Each of these variables was missing if the age or relationship codes for at least one roster member in a respondent's household was missing.

8.5 Imputation of Household Roster-Derived Variables

Although nine roster-derived variables were created from the edited roster, the missing values were imputed for only four of these variables: TOTPEOP, KID17, HH65, and FAMSKIP. The missing values in these variables were imputed using the UPMN technique 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 would be modeled. Each roster-derived variable was expected to be strongly related to the other three roster-derived variables. Hence, it was important to perform the imputations sequentially so that variables early in the series could be used as covariates for subsequent variables. The order in which the roster variables were imputed is shown in **Exhibit 17**.

Exhibit 17. Household Roster-Derived Variables (in Order of Imputation)

Roster Variable	Edited Variable	Imputed Variable
Total number of rostered people	TOTPEOP	IRHHSIZE
Total number of kids under 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

[1] FAMSKIP was set to 0 if the roster had relationship codes of 2, 3, 4, 5, 6, 8, 9, 10, 11, and 13 in **Exhibit 15**. FAMSKIP was set to 0 if no relationship codes were missing, and the roster had codes of 1, 7, 12, and/or 14.

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 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. 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 considered for the main model considered in the next section. The item response propensity model is described in greater detail in **Appendix B**.

8.5.3 Sequential Model Building

The variables TOTPEOP, KID17, and HH65 were assumed to have a Poisson distribution, and the parameters for the models were estimated using PROC GENMOD. The binary variable FAMSKIP was modeled using a weighted logistic regression. The covariates in each response propensity model were continuous age, age squared, gender, race/ethnicity, roster-derived variables earlier in the sequence, region, population density, percent Hispanic households in segment, percent of owner-occupied households in segment, and (for TOTPEOP only) number of people in the household eligible for interviewing (from the pre-interview screener). There were also predictors that consisted of one-way interactions of age with race/ethnicity, age with gender, race/ethnicity with gender, age squared with race/ethnicity, and age squared with gender. For the three older age groups, the additional covariates of marital status, education status, and employment status were also included as covariates.

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 D**.

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.

8.5.6 Constraints on Univariate Predictive Mean Neighborhoods

A univariate imputation was implemented on each variable within each age group after predicted values from the models had been determined. 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 logical constraints on the neighborhoods were sequentially based on the information already available in the roster, and 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 aged 18 or older, one with an age missing, and one with an age under 18, KID17 would be missing. Logically, however, at least one child under 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. HH65 was restricted within bounds in the same manner, using the variables IRHHSIZE, IRKID17 and the nonmissing ages in the roster.

Likeness constraints were also applied to the imputation of missing values in KID17, HH65, and FAMSKIP. A small delta (5 percent) could be considered a likeness constraint, which could be loosened by enlarging delta, or abandoning the neighborhood altogether and taking the donor with the closest predicted mean. 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), and FAMSKIP donors and recipients were required to have the same values for IRKID17 (the imputation-revised version of KID17). For KID17 and HH65, the household size likeness constraint was loosened after abandoning the neighborhood. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix G**.

References

Bowman, K.R., Penne, M.A., Westlake, M.J., & Singh, A.C. (1999, June). *1998 National Household Survey on Drug Abuse: Editing, imputation, and outlier weight evaluation report* (prepared for Substance Abuse and Mental Health Services Administration, Contract No. 283-97-9007, Deliverable No. 16; RTI/06908-082). Research Triangle Park, NC: RTI.

Bowman, K., Penne, M., Chromy, J., & Odom, D. (2002, May). *The 2000 National Household Survey on Drug Abuse: Sample design report (for inclusion in the 2000 methodological resource book)* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 10; RTI/07190.230.400). Research Triangle Park, NC: RTI.

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*, 376–382.

Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 197–201). Washington, DC: American Statistical Association.

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 American Statistical Association, Survey Research Methods Section* (pp. 598-603). Alexandria, VA: American Statistical Association.

Folsom, R.E., & Witt, M.B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. In *Proceedings of the American Statistical Association, Social Statistics Section* (pp. 428-433). Alexandria, VA: American Statistical Association.

Kroutil, L.A. (2002a, June). *2000 National Household Survey on Drug Abuse: General principles and procedures for editing drug use data in the 2000 NHSDA computer-assisted interview (for inclusion in the 2000 methodological resource book)* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.295). Research Triangle Park, NC: RTI.

Kroutil, L.A. (2002b, June). *2000 National Household Survey on Drug Abuse: Procedures for editing supplementary self-administered data in the 2000 NHSDA computer-assisted interview (for inclusion in the 2000 methodological resource book)* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.295). Research Triangle Park, NC: RTI.

Kroutil, L.A. (2002c, June). *2000 National Household Survey on Drug Abuse: Procedures for editing interviewer-administered data in the 2000 NHSDA computer-assisted interview (for inclusion in the 2000 methodological resource book)* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.295). Research Triangle Park, NC: RTI.

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

Manly, B. F. J. (1986). *Multivariate statistical methods: A primer*. London, England: Chapman and Hall.

Office of Applied Studies. (1999). *1998 National Household Survey on Drug Abuse: Methodological resource book*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2001, March). *Development of computer-assisted interviewing procedures for the 1999 National Household Survey on Drug Abuse* [On-line]. Available at http://www.samhsa.gov/oas/nhsda/CompAssistInterview/cover.htm [2001, June 20].

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 *1998 Social Statistics Section: Proceedings of the American Statistical Association* (pp. 744-749). Washington, DC: American Statistical Association.

Penne, M.A., & Witt, M.B. (1997). *1997 National Household Survey on Drug Abuse: Adjusting the sample design to allow for dwelling unit level estimation* (prepared for Substance Abuse and Mental Health Services Administration, Contract No. 283-96-0001). Research Triangle Park, NC: RTI.

RTI. (2001). *SUDAAN user's manual: Release 8.0*. Research Traingle Park, NS: Author.

RTI. (2002, July). *2000 National Household Survey on Drug Abuse specifications for programming (for inclusion in the 2000 methodological resource book)* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190). Research Triangle Park, NC: Author.

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

SAS Institute. (1999). *SAS/STAT user's guide: Version 8*. Cary, NC: Author.

Schafer, J.L. (1997). *Analysis of incomplete multivariate data*. Boca Raton, FL: Chapman and Hall/CRC.

Singh, A.C., & Mohl, C.A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology*, *22*(2),107-115.

Appendix A: Unweighted Hot-Deck Method of Imputation

Appendix A: Unweighted Hot-Deck Method of Imputation

A.1 Introduction

With the unweighted 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."

Two types of unweighted hot-deck imputation were used in the 2000 National Household Survey on Drug Abuse (NHSDA). The first method, the unweighted sequential hot deck, was the exclusive method of hot-deck imputation used for the 1991 to 1998 NHSDAs and the paper-and-pencil interviewing (PAPI) sample of the 1999 NHSDA. For 1999 and 2000 computer-assisted interviewing (CAI), unweighted sequential hot deck was only used in certain item imputations. As with the 1999 NHSDA's CAI, the 2000 NHSDA also used the second method, the unweighted random nearest neighbor hot deck (NNHD). These methods are discussed in the following sections. With both types of unweighted hot-deck imputation, the identity of the donors is tracked. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

A.2 Unweighted Sequential Hot Deck

The implementation of the unweighted sequential hot deck involved three basic steps, as described in the following sections.

A.2.1 Forming 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. In the main body of the report, these classes were defined by logical and likeness constraints, where classes defined by the likeness constraints could be collapsed if insufficient donors were available, and those defined by logical constraints could not be collapsed, due to the possibility of an inconsistency with preexisting nonmissing values that would result.

A.2.2 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 could be used 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 on. 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) changes each time the value of a variable changes. 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.2.3 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, and any subsequent nonrespondent encountered received the stored donor response creating the statistically imputed response. A starting value was needed if an item nonrespondent was the first record on a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value.

Note that 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.2.4 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 the NHSDA, the imputation donor was identified for every item being imputed. Then, by examining frequencies by imputation donor, one could see whether several nonrespondents were lining up next to one another in the sort. When this problem occurred, sort variables could be added, eliminated, or the order of the variables could be rearranged.

A.3 Unweighted Random Nearest Neighbor Hot Deck

As with the unweighted sequential hot deck, the unweighted random nearest neighbor hot deck (NNHD) can be implemented in three steps, the first of which is identical to the unweighted sequential hot deck.

A.3.1 Forming 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. In the main body of the report, these classes were defined by logical and likeness constraints, where classes defined by the likeness constraints could be collapsed if insufficient donors were available, and those defined by logical constraints could not be collapsed, due to the possibility of an inconsistency with preexisting nonmissing values that would result.

A.3.2 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, one could calculate the distance between the values of the recipient and potential donors for each of the auxiliary variables, then choose donors for the neighborhood such that the maximum of these distances was less than a certain value, referred to as "delta." This neighborhood could be 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 the NHSDA, 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 could be made relative by dividing by the predicted mean of the recipient, so that delta could represent a percentage.

A.3.3 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 whose base variable values would replace those of the recipient. The selection could be conducted as a simple random sample or could incorporate the weights of the potential donors.

Appendix B: Technical Details about the Generalized Exponential Model (GEM)

Appendix B: Technical Details about the Generalized Exponential Model (GEM)

B.1 Distance Function

Let $\Delta(\boldsymbol{w},\boldsymbol{d})$ denote the distance between the initial weights $\boldsymbol{d} = \{d_k : k \in s\}$ and the adjusted weights **w**. The distance function minimized under the generalized exponential model (GEM) subject to calibration constraints is given by

$$\Delta(w,d) = \Sigma_{k\in s} \frac{d_k}{A_k}\left\{(a_k - \ell_k)\log\frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k)\log\frac{u_k - a_k}{u_k - c_k}\right\} \tag{B1.1}$$

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 predictor variables (x_1, ..., x_p, say). Then the calibration constraints for the above minimization problem are

$$\Sigma_{k\in s}\, x_k\, d_k\, a_k = T_x \tag{B1.2}$$

The solution of the above minimization problem, if it exists, is given by a GEM with model parameters λ, viz,

$$a_k(\lambda) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp\{A_k \boldsymbol{x}_k'\lambda\}}{(u_k - c_k) + (c_k - \ell_k)\exp\{A_k \boldsymbol{x}_k'\lambda\}} \tag{B1.3}$$

Note that the number of parameters in GEM should be $\le n$, where n is the size of the sample s. This is also the dimension of vectors **d** and **w**. It follows from (B1.3) that

$$\ell_k < a_k < u_k,\ k = 1,..,n \tag{B1.4}$$

The usual Raking-ratio method (see e.g., Singh & Mohl, 1996) of weight adjustment is a special case of GEM by noting that for $\ell_k = 0,\ u_k = \infty,\ c_k = 1,\ k = 1,..,n$,

$$\Delta(w,d) = \sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k(a_k - 1) \quad \text{(B1.5)}$$

and $a_k(\lambda) = exp(\boldsymbol{x}_k'\lambda)$.

The logit method of Deville and Särndal (1992) is also a special case of GEM by setting $\ell_k = \ell,\ u_k = u,\ c_k = 1$ for all k. The new method was introduced by Folsom and Singh (2000). More details can be found there.

B.2 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 (B1.3) can be justified for all the three types of adjustment. For extreme value (ev) treatment via winsorization, denote the winsorized weights by $\{b_k\}$ where $b_k = d_k$ if d_k is not an outlier, and = med$\{d_k\} \pm 3*$ IQR if d_k is an outlier, where the quartiles for the weights are defined with respect to a suitable design-based stratum. Then with GEM for outlier treatment, $\ell_k = 1, c_k = c = 1 + \Sigma_{s**}(d_k - b_k)/\Sigma_{s*} d_k$ and $u_k = u > c$ can be chosen for nonoutliers, and the outliers are held fixed at their winsorized values, where s_* denotes the subsample of nonoutliers, and s_{**} the subsample of outliers.

For the nonresponse (nr) adjustment, the sample is divided as before in two parts, s_* the nonoutlier subsample, and s_{**} the outlier subsample. For nonoutliers, l_2 is set as $\ell_2 = 1, c_2 = \rho^{-1}, u_2 = u > \rho^{-1}$, where ρ is the overall response propensity; and for outliers with high weights, l_k is set as $\ell_k = \ell_1 m_k, c_k = m_k, u_k = u_1 m_k$, where $m_k = b_k/d_k$, and $\ell_1 < 1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, $1 < \ell_3 < \rho^{-1} = c_3 < u_3$ is set for outliers with low weights.

For the poststratification (ps) adjustment, l_k is set for nonoutliers as $\ell_k = \ell_2, c_k = c_2 = 1, u_k = u_2$, and for high outliers, $\ell_k = \ell_1 m_k, c_k = m_k, u_k = u_1 m_k$, and similarly for low outliers.

Notice that with GEM, one has the flexibility of specifying different bounds for different subsamples, as well as the lower bound (in the case of outlier and nr adjustments) can be made 1 by choosing the centre $c_k > 1$.

B.3 Newton-Raphson Steps

Let **X** denote the *n* x *p* matrix of predictor values, and for the νth iteration,

$$\Gamma_{\phi\nu} = \text{diag } (d_k \, \phi_k^{(\nu)}), \ \phi_k^{(o)} = 1$$

where

$$\phi_k^{(\nu)} = (u_k - a_k^{(\nu)}) \, (a_k^{(\nu)} - \ell_k) \, / \, (u_k - c_k) \, (c_k - \ell_k) \quad .$$

Then at the Newton-Rahpson iteration ν, the value of the p-vector λ is adjusted as

$$\lambda^{(\nu)} = \lambda^{(\nu-1)} + (X' \, \Gamma_{\phi, \nu-1} X)^{-1} \, (T_x - \hat{T}_x^{(\nu-1)}) \tag{B3.1}$$

where $\boldsymbol{\lambda}^{(0)} = \mathbf{1}$.

The convergence criterion is based on the Euclidean distance $\| \boldsymbol{T_x} - \hat{\boldsymbol{T}}_x^{(\nu)} \|$. At each iteration, it is checked whether it is decreasing or not. If not, then half-step is used in the iteration increment.

B.4 Scaled Constrained Exponential Model

In previous NHSDAs, constrained exponential models (CEM) were used for ps and scaled CEM for nr 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 nr adjustment, Folsom and Witt (1994) modified CEM estimating equations by a scaling factor (ρ^{-1}: 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, one can ensure 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

At RTI, a new approach was developed for the imputation of missing values in the computer-assisted interviewing (CAI) sample of the 1999 National Household Survey on Drug Abuse (NHSDA). This approach can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, it incorporates predictive means from models and the assignment of imputed values using neighborhoods determined by those predictive means.

C.2 Overview

C.2.1 Predictive Mean Neighborhoods, Derived from Combining Nearest Neighbor Hot Deck and Predictive Mean Matching

The new method, called predictive mean neighborhood (PMN), is a combination of two commonly used imputation methods: a non-model-based hot deck (nearest neighbor), and a modification of the model-assisted predictive mean matching (PMM) method of Rubin (1986). PMN enhances the PMM method in that it can be applied to both discrete and continuous variables either individually or jointly. PMN also enhances the nearest neighbor hot-deck (NNHD) method in that the distance function used to find neighbors is no longer ad hoc.

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, the recipient does not). A donor set deemed close to the recipient with respect to a number of covariates is used to select a donor at random. For the NHSDA, the set of covariates typically would include demographic variables as well as some other nonmissing drug use variables. To further ensure that a donor matches the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with drug use, such as age categories, can be used to restrict the set of donors. Furthermore, other restrictions involving outcome variables can be imposed on the neighborhood. Note that in NNHD, unlike 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 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 predictive mean for the recipient, obtained under a model, and the response variable outcomes for candidate donors. The respondent with the smallest distance is chosen as the donor. Unlike the NNHD, the donor is not randomly selected from a neighborhood. The advantages of PMM include the following:

- Model bias in the predictive mean can be minimized by using suitable covariates.

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

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

C.2.2 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

PMN is easily applicable to problems of both univariate and multivariate imputations. The need for univariate imputation arises when 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, while 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, 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 sheer number 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, none of the existing solutions take the survey design into account because of the obvious problem of specifying the probability distribution underlying survey data. However, in the application of the multivariate predictive mean neighborhood (MPMN) imputation to the 1999 NHSDA, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case) such that variables modeled earlier on in the hierarchy have a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for use of univariate techniques for fitting multivariate (but conditional) predictive means.

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

C.3 Outline and Description of Method

The procedure for implementing UPMN and MPMN entails six steps. Steps 2 through 5, and sometimes Step 6, are cycled through each of the drugs and drug use measures in the order determined by Step 1. Steps 4 and 5 (Steps 4 to 6 when applicable) could be considered a variant of a random NNHD.

C.3.1 Step 1: Hierarchy Definition

The first step is to determine the order in which variables are modeled, so that variables early in the hierarchy may be used for modeling the conditional predictive mean (i.e., they have the potential to be part of the set of covariates for variables later in the hierarchy). Note that not all variables in the hierarchy may be missing for a particular incomplete record. Nevertheless, models are to be developed for all the variables in a univariate fashion for reasons mentioned earlier. For example, in the drug modules in the CAI sample of the 1999 NHSDA, different drugs needed to be 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 were to be 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 the NHSDA, 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 be 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 are to be followed.

C.3.2 Step 2: Setup for Model Building and Hot-Deck Assignment

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

C.3.3 Step 3: Sequential Hierarchical Modeling

The model is to be built using the complete data respondents only, with weights adjusted for item nonresponse. For the CAI drug modules, lifetime usage indicators are to be modeled first because all other drug use indicators depend on an indication of lifetime use or nonuse. Once the hierarchy of drugs for lifetime usage has been determined, lifetime usage indicators for individual drugs can be modeled in a sequential fashion. The sequence used for the remaining combinations of drugs and drug use measures depends on what covariates are desired in the models and what variables are considered part of a multivariate set.

C.3.4 Step 4: Computation of Predictive Means and Delta Neighborhoods

Once the model has been fitted, the predictive means for item respondents and item nonrespondents are to be calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element out of that vector must be chosen, so that each respondent has exactly one predictive mean value[72] This predictive mean is the matching variable in a random NNHD. It can come directly from the model, it can be adjusted to account for the conditioning on the time period, or (if it is the predicted value based on a model with a transformed response variable) it can be back-transformed to the original units.

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

The pool of donors is to be further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a "logical constraint," is given by age at first crack use, which must not be younger than age at first cocaine use. Other constraints, called "likeness constraints," are 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 are restricted to be the same whenever possible, and the donor and recipient must come from States with similar usage patterns. A small value of delta could also be thought of as a likeness constraint. Whenever insufficient donors are available to meet the likeness constraints, including the preset small value of delta, the constraints are to be loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors are available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor is to be found using a sequential hot deck, where matching is to be done on the predictive mean. (Even though weights would not be used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process because weighting would already have been incorporated in the calculation of the predicted mean.)

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

---

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

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 is to be chosen for each item nonrespondent. If only one response variable is to be imputed, the assignment step is just a simple replacement of a missing value by the value of the donor. It is possible, however, that a donated quantity is a function of the final imputed value. For example, for 12-month frequency of drug use, because donors and recipients could potentially have a different maximum possible number of days in the year that they could have used a substance, the observed proportion of total period is donated rather than the observed 12-month frequency, where the "total period" could range up to a year. In the assignment step, the donor's proportion of total period is to be 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 is multivariate if several response variables are associated with a single predictive mean, provided more than one of those response variables is missing. In that case, all of the missing values are to be imputed using the same donor. If there is more than one response variable associated with a single predictive mean, but not all of them are missing, only the missing values are to be replaced by those of the donor. The resulting imputed values are provisional if a multivariate neighborhood (MPMN) step is called for; otherwise, these values are final.[73]

> *If the variables for which Steps 2 to 5 have been completed are part of a complete multivariate set for which MPMN is to be applied, Step 6 is the next step in the process. If the variables for which Steps 2 to 5 are completed are not part of a complete multivariate set, and other variables are still to be imputed, Step 2 is the next step. Otherwise, the process is finished.*

C.3.6 Step 6: Determination of Multivariate Predictive Mean Neighborhood and Assignment of Imputed Values

With MPMN, the neighborhood is defined based on a vector of predictive means rather than from a single predictive mean as in the univariate case. This vector may encompass a subvector of predictive means from a single categorical model (as with a polytomous logit model), in addition to scalar predictive means from any number of models with continuous response variables. For each item nonrespondent, a distance is to be calculated between the elements of this vector of predictive means where the observed values are missing, and the

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

corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predictive means that were calculated on the basis of past year and past month users are to be adjusted to account for the probability that a respondent is a past year user or a past month user. For example, in the CAI sample of the 1999 NHSDA, the full predictive mean vector for alcohol 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)[74];

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 depends on the missingness pattern of that item nonrespondent.[75] Moreover, if partial information is available on the recency of use, the predictive means is to be adjusted to account for that knowledge. For example, if a particular item nonrespondent was known to be a past year alcohol user and his 12-month frequency was known, the elements above for which differences would be calculated would be element #1 conditioned on past year use, and #4 and #5. That is,

$$P\ (\textit{Past month alcohol user} \mid \textit{Lifetime alcohol user}) \div P\ (\textit{Past year alcohol user} \mid \textit{Lifetime alcohol user}),$$

[74] For the 12-month frequency, 30-day frequency, and 30-day binge frequency, the models are fit using logits. These logits are 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 assumes that the probability of use on each day in the year is equal. This, of course, is not true. However, the violation of this assumption does not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it does allow a predicted mean to be made conditional on what is known.

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

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*).[76]

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

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

where μ_R refers to the predictive mean (sub-)vector for a given item respondent, and μ_{NR} is the predictive mean (sub-)vector for a given item nonrespondent. The matrix Σ is the variance-covariance matrix of the predictive means, calculated using the subvector of predictive means associated with each missingness pattern, using complete data respondents within each age group and (where applicable) State rank group. The Mahalanobis distance is only to be calculated for those respondents who meet the delta constraint. The neighborhood is determined by selecting the *k* smallest Mahalanobis distances within this subset of item respondents for a given item nonrespondent.

If some of the variables in the response vector are not missing, only those that are missing are to be replaced. However, logical constraints must be placed on the multivariate neighborhood, so that imputed values are consistent with preexisting nonmissing values. For example, if a respondent is missing a 30-day frequency, but his or her nonmissing 12-month frequency is 350, a donor cannot have a 30-day frequency smaller than 350 - 335, or 15. If the number of respondents in the univariate subset who meet the logical constraints imposed upon the multivariate neighborhood is fewer than *k* but greater than 0, all the respondents in the resulting subset are to be selected for the neighborhood. Finally, if there are no respondents within the univariate subset who meet the logical constraints imposed by the multivariate neighborhood, the *k* smallest Mahalanobis distances who meet the logical constraints among all candidate donors for a given item nonrespondent are to be selected for the neighborhood. In addition to the multivariate delta, likeness constraints are used to make the donors in the

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

neighborhood as much like the recipient as possible. These can be loosened if insufficient donors are available. Finally, as with the univariate neighborhood, an unweighted sequential hot deck is to be used as a last resort if insufficient donors are available who meet the logical constraints and the loosest set of likeness constraints allowable.

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

C.4 Comparison of PMN with Other Available 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, that 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.

- **The problem of sparse neighborhoods is considerably reduced, which makes it easier to implement restrictions on the donor set.** Because the distance function is defined as a continuous function of the predictive 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 have been 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.

- **The correlations across response variables is accounted for by making the imputation multivariate.**

- **The choice of donor can be made random by choosing delta large enough such 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 has mean zero to 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, one can objectively assign differential weights to different elements of the predictive mean vector depending on the variability of predictive means in the dataset via the Mahalanobis distance.

As noted earlier, the PMN method has some similarity with the predictive mean matching method of Rubin (1986) except that, for the donor records, the observed variable value and not the predictive 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-Origin Group Alpha Codes

Appendix D: Race and Hispanic-Origin Group Alpha Codes

D.1 Introduction

To reduce the amount of statistical imputation necessary to create the imputation-revised race and Hispanic-origin variables for the computer-assisted interviewing (CAI) sample of the 2000 National Household Survey on Drug Abuse (NHSDA), the race and Hispanic-origin group alpha-specify dictionaries used in prior NHSDAs were expanded, as were the procedures used to assign the large number of other-specify responses to the categories used for these variables. As discussed in **Chapter 4**, many respondents provided a race in the alpha-specify response to the Hispanic-origin 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, using an expanded dictionary, in order to assign race and Hispanic-origin values to respondents based on alpha-specify responses.

D.2 Race

In a change from the paper-and-pencil interviewing (PAPI) questionnaire, respondents were permitted to select more than one race. There also was a follow-up question asking respondents who selected multiple races in the first question to select from among those chosen the single race that best described them. As in past years, respondents had an opportunity to specify a race not included in the question by responding "other," either as the sole race chosen in the first question, or as the race that best described them if "other" was among multiple races chosen in the first question. The race questions used in 2000 are as follows:

QD05: Which of these groups describes you? Just give me the number or numbers from the card.

1	White	7	Filipino
2	Black/African American	8	Japanese
3	American Indian or Alaska Native	9	Asian Indian
4	Native Hawaiian	10	Korean
5	Other Pacific Islander	11	Vietnamese
6	Chinese	12	Other Asian
		13	Other (Specify)

QD06: Which **one** of these groups, that is [races chosen in QD05], **best** describes you?

(Choose from among responses to QD05)

D.2.1 Race Alpha Responses

The other-specify responses were examined when (a) "other" was selected as a race in QD05,[77] or (b) no race was given in response to QD05, but a race category was given as an other-specify response to the Hispanic-origin group question (QD04). In such cases, if a respondent provided a valid alpha-specify response when asked, that response was used in order to assign a value of EDRACE, the base variable for imputing IRRACE (see **Chapter 4**), as follows:

1. The following other-specify responses were classified as "black": black (including part black), African American, Haitian, Caribbean Creole, Dominican (not Dominican Republic).

2. The following other-specify responses were classified as "Asian": Native Hawaiian, Other Pacific Islander, Chinese, Filipino, Japanese, Asian Indian, Korean, Vietnamese, Other Asian (including Iranian, Kurdish, Afghanistani), Asian nonspecific, and Guamanian. In addition, if a respondent indicated that he or she was a mix of any of the above Asian categories and some other race, other than black, or that he or she was partly of Hispanic origin and partly Asian (by indicating any of the above Asian categories), the respondent was classified as "Asian."

3. The following other-specify responses were classified as "American Indian": American Indian or Alaska Native (including mestizo) or part American Indian and part any other race except black or Asian. Also, any respondent indicating that he or she was part Hispanic and part American Indian was classified as "American Indian."

4. The following other-specify responses were classified as "white": white, North African, Arabic, Turkish, Armenian, Jewish, Middle Eastern/Israeli, Canadian, or part-Hispanic and part-white.

5. If a respondent indicated only an Hispanic-origin group in response to the race other-specify question, he or she was assigned to groups for restricted imputation of race. That is, race was statistically imputed for such respondents, using as donors only those respondents of the same Hispanic-origin group who gave a valid race response. The groups for restricted imputation were Hispanic nonspecific, Mexicans, Puerto Ricans, Cubans, Central or South Americans, Mexicans and Puerto Ricans combined,

[77] For the four-level variable IRRACE, this is relevant in two cases: (i) "other" was selected as the only race in QD05, or (ii) "other" was one of the multiple races selected in QD05 and was chosen as the "most descriptive" race in QD06. For the 15-level variable NEWRACE1, it is relevant for all cases where "other" was selected as one of several races in QD05, regardless of the race selected in QD06.

Mexicans and Central or South Americans combined, Mexicans and Cubans combined, Puerto Ricans and Central or South Americans combined, Puerto Ricans and Cubans combined, and Cubans and Central or South Americans combined.

6. For certain countries of origin given in the other-specify responses, race was randomly assigned using Census data for those countries. In many cases, a small percentage of respondents from a given country were left to be statistically imputed. The following is a list of the countries treated in this way and the percentages assigned to each race:[78]

   - Dominican Republic: 84 percent black, 16 percent white, 0 percent statistically imputed;

   - Caribbean and West Indies: 80 percent black, 14 percent Asian, 6 percent statistically imputed;

   - Belize: 55 percent American Indian, 37 percent black, 8 percent statistically imputed;

   - Guyana: 51 percent Asian, 43 percent black, 6 percent statistically imputed;

   - Suriname: 52 percent Asian, 31 percent black, 17 percent statistically imputed;

   - Trinidad and Tobago: 57 percent black, 40 percent Asian, 3 percent statistically imputed;

   - Jamaica: 91 percent black, 9 percent statistically imputed;

   - Bahamas and Virgin Islands: 85 percent black, 15 percent white, 0 percent statistically imputed;

   - Western Europe, including Spain and Portugal: 95 percent white, 5 percent statistically imputed;

   - New Zealand: 88 percent white, 9 percent black, 3 percent statistically imputed;

   - South Africa: 84 percent black, 13 percent white, 3 percent Asian, 0 percent statistically imputed; and

[78] Note that these are the percentages used to randomly assign respondents to races although the distribution of assigned races in the sample does not match these exactly. Also note that if 0 percent are statistically imputed, no respondents are assigned to the races that are not listed.

- Australia: 95 percent white, 4 percent Asian, 1 percent black, 0 percent statistically imputed.

D.2.2 Assigning a Race When Multiple Races Were Selected

As stated earlier, respondents were allowed to select more than one race in QD05 although they were asked to give the race that best represented them in QD06. Not all respondents who entered multiple races indicated a single race in QD06. In the imputation revised variable called IRRACE, only four races were given, and no category was available for multiple race. Hence, a decision rule had to be in place to determine which of the multiple races chosen would describe respondents who did not select a single race in QD05 or QD06. The priority rule in place was the same as that used in past years. That is, if a respondent indicated black/African American among any of his or her races, he or she was considered black/African American. Otherwise, if a respondent indicated any of the Asian categories as his or her race, he or she was considered Asian. If a respondent indicated neither black/African American nor any of the Asian categories, but indicated Native American as one of his or her races, the respondent was considered Native American. Finally, white respondents were those who only indicated "white" and no other race. This priority rule was not necessary with the recodes NEWRACE1 and NEWRACE2 because a separate category was created specially for respondents who indicated more than one race, regardless of whether they indicated a single race in QD06.

D.2.3 Race Dictionary Codes

Codes were assigned to respondents based either on their response to the first 12 categories of QD05 (codes 1 to 12), or on their race alpha-specify responses (codes 21 to 985). Codes 21 to 32 are equivalent to codes 1 to 12, except that the race identification was obtained from the alpha-specify responses. The values of EDRACE were obtained using these codes (see **Section D.2.2**), which are presented below:

1 White
2 Black/African American
3 American Indian or Alaska Native
4 Native Hawaiian
5 Other Pacific Islander
6 Chinese
7 Filipino
8 Japanese
9 Asian Indian
10 Korean
11 Vietnamese
12 Other Asian
21 White (includes Arab, Turkish, Armenian, Jewish)
22 Black/African American (includes Haiti, St. Vincent, Dominica)
23 American Indian or Alaska Native (includes mestizo)
24 Native Hawaiian
25 Other Pacific Islander
26 Chinese
27 Filipino
28 Japanese
29 Asian Indian

30 Korean
31 Vietnamese
32 Other Asian (includes Iran, Kurd, Afghan)
33 Asian nonspecific
34 Guamanian
41 Hispanic (nonspecific, race not given)
42 Mexican
43 Puerto Rican
44 Central or South American (excludes Belize/Guyana/Suriname)
45 Cuban
46 Dominican Republic (Santo Domingo)
47 Dominica (Roseau)
48 Dominican (Dominican Republic vs. Dominica not clear)
49 Caribbean/West Indies
50 Belize
51 Guyana
52 Suriname
53 Trinidad and Tobago
54 Jamaica
55 Virgin Islands (St. Thomas, St. Croix), Bahamas
80 United Kingdom
81 Portugal/European Spanish
82 Spanish, maybe European
83 Other Western Europe
84 Middle East/Israel/North Africa
85 Canada
86 New Zealand
87 South Africa
88 Australia
101 Part Hispanic, part white
102 Part Hispanic, part black
103 Part Hispanic, part American Indian
104 Part Hispanic, part Asian
105 Part Hispanic, part black, part white
106 Part "Spanish," part black
107 Part "Spanish," part Indian
108 Part "Spanish," part Asian
121 Mexican and Puerto Rican
122 Mexican and Central or South American
123 Mexican and Cuban
124 Puerto Rican and Central or South American
125 Puerto Rican and Cuban
126 Cuban and Central or South American
127 Mexican and Jamaican
128 Puerto Rican and Jamaican
129 Central or South American and Jamaican
130 Cuban and Jamaican
131 Dominican and Mexican
132 Dominican and Puerto Rican
133 Dominican and Central or South American
134 Dominican and Cuban
135 Mexican and European
136 Puerto Rico and European
137 Central or South American and European
138 Cuban and European
139 Trinidad and Mexican
140 Trinidad and Puerto Rican
141 Trinidad and Central or South American
142 Trinidad and Cuban
143 Mexican and Asian
144 Puerto Rican and Asian
145 Central or South American and Asian
201 Biracial (nonspecific)
202 White and black
203 White and American Indian
204 White and Native Hawaiian
205 White and Other Pacific Islander
206 White and Chinese
207 White and Filipino
208 White and Japanese
209 White and Asian Indian
210 White and Korean
211 White and Vietnamese
212 White and Other Asian
213 White and Asian (nonspecific)
223 Black and American Indian

224 Black and Native Hawaiian
225 Black and Other Pacific Islander
226 Black and Chinese
227 Black and Filipino
228 Black and Japanese
229 Black and Asian Indian
230 Black and Korean
231 Black and Vietnamese
232 Black and Other Asian
233 Black and Asian (nonspecific)
244 American Indian and Native Hawaiian
245 American Indian and Other Pacific Islander
246 American Indian and Chinese
247 American Indian and Filipino
248 American Indian and Japanese
249 American Indian and Asian Indian
250 American Indian and Korean
251 American Indian and Vietnamese
252 American Indian and Other Asian
253 American Indian and Asian (nonspecific)
265 Native Hawaiian and Other Pacific Islander
266 Native Hawaiian and Chinese
267 Native Hawaiian and Filipino
268 Native Hawaiian and Japanese
269 Native Hawaiian and Asian Indian
270 Native Hawaiian and Korean
271 Native Hawaiian and Vietnamese
272 Native Hawaiian and Other Asian
273 Native Hawaiian and Asian (nonspecific)
286 Other Pacific Islander and Chinese
287 Other Pacific Islander and Filipino
288 Other Pacific Islander and Japanese
289 Other Pacific Islander and Asian Indian
290 Other Pacific Islander and Korean
291 Other Pacific Islander and Vietnamese
292 Other Pacific Islander and Other Asian
293 Other Pacific Islander and Asian (nonspecific)
307 Chinese and Filipino
308 Chinese and Japanese
309 Chinese and Asian Indian
310 Chinese and Korean
311 Chinese and Vietnamese
312 Chinese and Other Asian
328 Filipino and Japanese
329 Filipino and Asian Indian
330 Filipino and Korean
331 Filipino and Vietnamese
332 Filipino and Other Asian
349 Japanese and Asian Indian
350 Japanese and Korean
351 Japanese and Vietnamese
352 Japanese and Other Asian
360 Asian Indian and Korean
361 Asian Indian and Vietnamese
362 Asian Indian and Other Asian
371 Korean and Vietnamese
372 Korean and Other Asian
382 Vietnamese and Other Asian
401 White, black, American Indian
402 White, black, Native Hawaiian
403 White, black, Other Pacific Islander
404 White, black, Chinese
405 White, black, Filipino
406 White, black, Japanese
407 White, black, Asian Indian
408 White, black, Korean
409 White, black, Vietnamese
410 White, black, Other Asian
411 White, black, Asian (nonspecific)
420 White, black, Hispanic
421 White, American Indian, Hispanic

900 Mixed/Mezclado 985 Bad data

D.3 Hispanicity

As with the race questions, Hispanic respondents[79] had the opportunity to specify a Hispanic-origin group by responding "other" to the Hispanic-origin group question (QD04). Also, unlike in the PAPI questionnaire, respondents were permitted to select multiple Hispanic-origin groups in response to QD04. However, unlike with the CAI race questions, there was no follow-up question asking respondents to choose a single group from among multiple groups chosen. The Hispanic-origin group question is as follows.

QD04: Which of these 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
5 Other (Specify)

D.3.1 Hispanic-Origin Group Alpha Responses

The other-specify responses were examined when (a) "other" was the only Hispanic-origin group selected in QD04, or (b) no Hispanic-origin group was given in response to QD04, but a Hispanic-origin group was given as an other-specify response to the race question (QD05). In such cases, if a respondent provided a valid alpha-specify response when asked, that response was used in order to assign a value of EDQD04, the base variable for imputing IRHOGRP/IRHOGRP3 (see **Chapter 4**), as follows:

1. The following other-specify responses were classified as "Mexican": Mexican (including part Mexican), Mexican American, Mexicano, Chicano.

2. The following other-specify responses were classified as "Cuban": Cuban, Cuban American, and part Cuban and part any other Hispanic-origin group except Mexican.

3. The following other-specify responses were classified as "Puerto Rican": Puerto Rican and part Puerto Rican and part Central or South American.

[79] For the purposes of the CAI instrument question-routing, Hispanic respondents were identified by their response to question QD03: "Are you of Hispanic, Latino, or Spanish origin or descent?"

4. Respondents who gave an other-specify response of "Central or South American" were classified into that category.

5. The following other-specify responses were classified as "Caribbean Islander": Hispanic Caribbean Islander (includes Dominican Republic and Santo Domingo), Dominican (where Dominica vs. Dominican Republic unclear).

6. If a respondent indicated only a race in response to the Hispanic-origin group other-specify question, he or she was assigned to a group for restricted imputation of Hispanic-origin group. That is, an Hispanic-origin group was statistically imputed for such respondents, using as donors only those respondents of the same race who gave a valid Hispanic-origin group response. The groups used for restricted imputation were whites, blacks, American Indians, Asians, and blacks and whites combined.

D.3.2 Hispanic-Origin Group Dictionary Codes

Codes were assigned to respondents based either on their response to the first four categories of QD05 (codes 1 to 4), or on their Hispanicity alpha-specify responses (codes 11 to 85). Codes 11 to 14 are equivalent to codes 1 to 4, except that the race identification was obtained from the alpha-specify responses. The values of EDQD04 were obtained using these codes (see **Section D.2.2**), which are presented below:

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican
3 Central or South American
4 Cuban/Cuban American
11 Mexican/Mexican American/Mexicano/Chicano
12 Puerto Rican
13 Central or South American
14 Cuban/Cuban American
21 Mexican/Puerto Rican
22 Mexican/Central or South American
23 Mexican/Cuban
24 Puerto Rican/Central or South American
25 Puerto Rican/Cuban
26 Central or South American/Cuban
31 Hispanic Caribbean (includes Dominican Republic, Santo Domingo)
32 Belize (formerly British Honduras)
33 Dominican (Dominica vs. Dominican Republic unclear)
34 Other Caribbean, possibly Hispanic
35 Portugal/European Spanish/Basque/Canary/Cape Verde
36 "Spanish," non-European versus European unclear
37 Philippines/Guam
38 Spanish Filipino or Spanish Guamanian
50 (All) Hispanic, white, no other information
51 (All) Hispanic, black, no other information
52 (All) Hispanic, American Indian, no other information

53 (All) Hispanic, Asian, no other information
54 (All) Hispanic, no other information
60 Part Hispanic, part white
61 Part Hispanic, part black
62 Part Hispanic, part American Indian
63 Part Hispanic, part Asian
64 Part Hispanic, part black, part white
65 Part "Spanish," part black
66 Part "Spanish," part Indian
67 Part "Spanish," part Asian
70 Other possibly Hispanic (white)
71 Other possibly Hispanic (black)
72 Other possibly Hispanic (American Indian)
73 Other possibly Hispanic (Asian)
74 Other possibly Hispanic (multiracial)
75 Other possibly Hispanic (New Mexico)
76 Other possibly Hispanic (Texas)
77 Other possibly Hispanic (California)
80 Other definitely not Hispanic (includes Dominica)
85 Bad Data / "Mixed" / "Mezclado"

Appendix E: Employment Status Alpha Codes for CAI

Appendix E: Employment Status Alpha Codes for CAI

E.1 Introduction

As discussed in **Chapter 5**, the employment status questions in the computer-assisted interviewing (CAI) questionnaire of the 1999 and 2000 National Household Survey on Drug Abuse (NHSDA) were quite different from the paper-and-pencil interviewing (PAPI) questionnaire used in prior NHSDAs. The questions appeared in the noncore section of questionnaire, and a respondent's current employment status was determined from responses to a series of questions (QD26 to QD39) regarding work patterns in the past week and past 12 months.

E.2 Questions Regarding Respondents' Reasons for Not Working in the Past Week

As part of this series of questions, respondents who indicated that they did not work in the week prior to the interview were asked their reason for not working. Respondents who indicated that although they did not work in the past week, but did have a job or business, were routed to question QD30, which asked why they did not work during that week. Respondents who indicated that they did not work in the past week and did not have a job or business were routed to QD31, which asked why they did not have a job or business during that week. Both of these questions had "Some Other Reason" as a possible response, and respondents who chose this answer were asked to specify the reason. Questions QD30 and QD31 are listed below.

QD30: Please look at this card and tell me which of these reasons **best** describes why you did not work last week. Just give me the number.

1 On vacation/Sick/Furlough/Strike/Other temporary absence
2 On layoff and **not** looking for work
3 On layoff and looking for work
4 Waiting to report to a new job
5 Self-employed and did not have any business last week
6 Going to school/training
7 Some other reason

QD31: Please look at this card and tell me which one of these reasons **best** describes why you did not have a job or business last week. Just give me the number.

1 Unemployed or on layoff and looking for work
2 On layoff and **not** looking for work

3 Keeping house full time
4 Going to school/training
5 Retired
6 Disabled for work
7 Some other reason

E.3 Not Working Alpha Responses and Dictionary Codes

If the response given to the interviewer to either of these questions was a "7," the interviewer could type in the respondent's answer(s) that did not match the first six choices. To map these responses to the edited variable JOBSTAT, responses were coded using employment status dictionaries, one dictionary for each question. QD30SP was the other-specify question for QD30. Individuals were routed there if they said that they had a job or business, and the reason they did not work in the past week could not be described by one of the six choices. The codes for QD30SP (21 to 62) follow:

21 On vacation/sick/furlough/strike/other temporary absence
22 On layoff and not looking for work
23 On layoff and looking for work
24 Waiting to report to a new job
25 Self-employed, no business last week
26 Going to school/training
42 Respondent has a part-time job
43 Recently unemployed, no further information
44 Seasonal work
45 Employed, no further information
46 Retired
47 Homemaker
48 Disabled or in ill health, work status unclear
49 Not scheduled or not needed
50 Babysitting
51 Didn't want to work
52 Volunteer, work stat unclear
53 Not eligible to work
54 Works during school year only (e.g., teacher)
55 Temporary job, work status unclear
56 Active in other activities, work status unclear
57 Didn't need to work
58 Incarcerated, work status unclear
59 Semi-retired, work status unclear
60 Summer camp, work status unclear
61 On summer break, student or teacher unclear
62 Daycare unavailable, work status unclear

QD31SP was the other-specify question for QD31. Individuals were routed there if they said that they did not have a job or business or if they did not answer the question regarding whether they had a job or business, and the reason they did not have a job or business could not be described by one of the six choices. The codes for QD31SP (21 to 84) follow:

21 Unemployed or on layoff and looking for work
22 On layoff and not looking for work
23 Keeping house full time
24 Going to school/training
25 Retired
26 Disabled for work
41 Respondent has a full-time job
42 Respondent has a part-time job
43 Temporary absence from work
44 Unemployed, no further information
45 Doesn't want to work/not interested in working
46 Doesn't need to work
47 Not eligible to work (too young/no work permit)
48 Married/pregnant/gave birth/divorce, work status unclear
49 Recently moved/new resident, work status unclear
50 Waiting to report to new job
51 Volunteer work, no other information
52 Seasonal work
53 Active in sports, work status unclear
54 Waiting to start school
55 Caring for disabled/ill/elderly relative
56 Not working due to location/no transportation
57 Helping parents/responsibilities at home
58 No permission to work from parent or guardian
59 Finished or quit school, not working
60 Student/youth, currently looking for work
61 Injured/ill, unclear whether disabled for work
62 Babysitting
63 Substance abuse issues
64 Do not work outside religious community/commune
65 Doesn't work/never worked, reason unspecified
67 Incarcerated/criminal record
68 Still deciding what to do
69 Not scheduled or not needed
70 Do not earn enough money
71 Active in other activities, work status unclear
73 Has temporary job, work status unclear
74 Illiterate/learning disability/mental barrier/language barrier
75 Working from/around home/work status unclear
76 Lawsuit, advised against working
77 Religious mission/work, paid or unpaid unclear
78 Working, full or part-time unclear
79 Works during school year only (e.g., teacher)
80 Starting new business
81 On summer break, student or teacher unclear
82 Age constraint, eligibility to work unclear
83 Daycare unavailable, work status unclear
84 Summer camp, work status unclear

Based on responses to the employment status questions, including the alpha-specify responses provided in QD30SP and QD31SP, a logically edited employment status variable (JOBSTAT) was created and used as a base variable for creating the final imputed employment status variables EMPSTAT3 and EMPSTT3R (see **Chapter 5** and editing documentation). JOBSTAT had many more categories than the final variables EMPSTAT3 and EMPSTT3R.

Respondents' JOBSTAT values were recoded into the categories of employment status in the final variable as follows. Note that all respondents aged 12 to 17 were assigned to a single category of the final employment status variable.

1. The following JOBSTAT categories were classified as *"Employed full-time"*: worked full-time last week, work full-time during school year, has full-time job and reason for not working unknown. Furthermore, respondents who indicated that they had a job, but were out during the last week, **and** that they usually work 35 or more hours per week were classified as "Employed full-time" if they had one of the following JOBSTAT values: has job but out (vacation/sick/temporary absence), has job but out (waiting to report to new job), has job but out (self-employed, no business), has job but out (in school/training), not scheduled/temporary/on-call worker, babysitter, has job and did not want to work last week, has a job during school year (no further information), has a job (no further information).

2. The following JOBSTAT categories were classified as *"Employed part-time"*: worked part-time last week, has part-time job and reason for not working unknown. Furthermore, respondents who indicated that they had a job, but were out during the last week, **and** that they usually work fewer than 35 hours per week were classified as "Employed part-time" if they had one of the following JOBSTAT values: has job but out (vacation/sick/temporary absence), has job but out (waiting to report to new job), has job but out (self-employed, no business), has job but out (in school/training), not scheduled/temporary/on-call worker, babysitter, has job and did not want to work last week, has a job during school year (no further information), has a job (no further information).

3. The following JOBSTAT categories were classified as *"Unemployed"*: no job (unemployed/on layoff and looking for work), no job (on layoff and not looking for work), unemployed (no further information).

4. The following JOBSTAT categories were classified as *"Other"*: has job but out (on layoff and looking for work), has job but out (on layoff and not looking for work), no job (keeping house full time/in school or training/retired/disabled for work/family responsibilities/starting or finishing school/substance abuse issues/criminal record/income restrictions/language or literacy problems/learning disability/legal issues), seasonal worker, volunteer worker, does not need to work, does not want to work, cannot work (reason unspecified), not eligible/allowed to work, student or youth (looking for work), doesn't work/never worked (reason unspecified), other (not in labor force).

5. If all that could be determined from a respondent's answers is that he or she had a job, the final employment status classification was assigned via

statistical imputation, but donors were restricted to respondents with valid employment status responses who were known to be either full-time or part-time employed. This restricted imputation was used for respondents who indicated that they had a job, but were out during the last week, **and** did not indicate clearly whether they usually work 35 or more hours per week **and** had any of the following JOBSTAT values: has job but out (vacation/sick/temporary absence), has job but out (waiting to report to new job), has job but out (self-employed, no business), has job but out (in school/training), not scheduled/temporary/on-call worker, babysitter, has job and did not want to work last week, has a job during school year (no further information), has a job (no further information).

The other final employment status variables, EMPSTAT4 and EMPSTATY, did not use the alpha-specify responses.

Appendix F: Model Summaries

Appendix F: Model Summaries

F.1 Introduction

The following tables list the covariates used in all models run during the 2000 imputation procedures. For each variable or set of variables to which a predictive mean neighborhood (PMN) method was applied, two models were run: one to adjust the weights for item nonresponse (response propensity models), and a second to calculate predictive means. Imputation was usually done separately among age groups, so most of the tables are for only one age group.

Section F.2 deals with the demographic variables; **Section F.3** deals with the drug variables; **Section F.4** deals with the health insurance variables; **Section F.5** deals with the income variables; and **Section F.6** deals with the household composition variables.

F.2 Demographic Variables

Exhibit F.1 Model Summaries (Apply to All Three Age Groups)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Race	Census Region; Household Type; Age; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households	Census Region; Household Type; Age; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households
Hispanic Origin	Census Region; Imputation Revised Race; Age; Age^2; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households	Census Region; Imputation Revised Race; Household Type; Age; Age^2; Age^3; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households
Marital Status	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Population Density; Age; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Age*Gender	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Population Density; Age; Age^2; Age^3; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Age*Gender
Hispanic Group	Census Region; Imputation Revised Race; Gender; Age; Age^2; Age^3; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Age*Gender; Age^2*Gender	Census Region; Imputation Revised Race; Gender; Age; Age^2; Age^3; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Age*Gender; Age^2*Gender

F.3 Drug Variables

Exhibit F.2 Cigarettes: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Census Region
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Age; Race; Gender; Gender*Race; Age*Race; Census Region; MSA; State Rank; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency
Age at First Daily Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarettes' Age at First Use

Exhibit F.3 Cigarettes: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Marital Status; Education; Employment Status; Census Region
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Age; Race; Gender; State Rank; Age*Race; Gender*Race; Age*Gender; Census Region; Marital Status; Education; Employment Status; MSA; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency
Age at First Daily Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarettes' Age at First Use

Exhibit F.4 Cigarettes: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	Age; Race; Gender; Age^2; Age^3; Age*Gender; Gender*Race; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3;Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Age; Race; Gender; State Rank; Age*Race; Gender*Race; Age*Gender; Census Region; Marital Status; Education; Employment Status; MSA; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Lifetime Indicators of Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Age; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency
Age at First Daily Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarettes' Age at First Use

Exhibit F.5 Cigars: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco Lifetime Indicator; Gender*Race; Age*Gender; Census region; MSA; State Rank
Recency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency;	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, and Smokeless Age at First Use

Exhibit F.6 Cigars: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco Lifetime Indicator; Age*Gender; Gender*Race; Census Region; MSA; State Rank; Education; Employment Status
Recency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency	Age; Race; Gender; Age^2; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, and Smokeless Age at First Use

Exhibit F.7 Cigars: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Census Region;Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco Lifetime Indicator; Age*Race; Age*Gender; Gender*Race; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Age; Race; Gender; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	Age; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Alcohol, Pipe, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette and Smokeless Tobacco Recency; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarettes' 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, and Smokeless Age at First Use

Exhibit F.8 Pipes: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco and Cigar Lifetime Indicators; Age*Race; Age*Gender; Race*Gender; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; State Rank
Recency	Race; Gender; MSA; Census Region; Imputation revised Cigarette Recency	Age; Gender; Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Analgesics, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Exhibit F.9 Pipes: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco and Cigar Lifetime Indicators; Age*Gender; Gender*Race; Marital Status; Education; Employment Status; MSA; Census Region; State Rank
Recency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency	Age; Gender; Race; Gender*Race ; Marital Status; Education; Employment Status; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigar, Alcohol, Inhalants, Marijuana, Hallucinogens, Analgesics, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Exhibit F.10 Pipes: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Smokeless Tobacco and Cigar Lifetime Indicators; Age*Gender; Race*Gender; Marital Status; Education; Employment Status; MSA; State Rank; Census Region
Recency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency	Age; Age^2; Age^3; Race; Gender; Age*Race; Age*Gender; Race*Gender; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime indicators of Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

Exhibit F.11 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	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Age*Race; Age*Gender; Gender*Race; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; State Rank; Census Region
Recency	<u>Smokeless Tobacco</u>: Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin <u>Chewing Tobacco</u>: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin <u>Snuff</u>: Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	<u>Smokeless Tobacco:</u> Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin <u>Chewing Tobacco:</u> Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin <u>Snuff:</u> Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A

Exhibit F.11 (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives Snuff: Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Chewing Tobacco: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives Snuff: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Chewing Tobacco and Snuff 30-Day Frequency; Imputation-Revised Cigarette and Cigarette Daily at First Use

Exhibit F.12 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	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Age*Race; Age*Gender; Gender*Race; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; Marital Status; Education; Employment Status; State Rank; Census Region
Recency	Smokeless Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Smokeless Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A

Exhibit F.12 (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Race; Gender; Census Region; MSA; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Pain Relievers, Heroin Snuff: Race; Gender; Census Region; MSA; Lifetime Indicators of Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Tranquilizers, Stimulants, and Sedatives	Chewing Tobacco: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives Snuff: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Chewing Tobacco and Snuff 30-Day Frequency; Imputation-Revised Cigarette and Cigarette Daily at First Use

Exhibit F.13 Smokeless Tobacco (Chewing Tobacco and Snuff): 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Age*Gender; Gender*Race; MSA; Marital Status; Education; Employment Status; State Rank; Census Region
Recency	Smokeless Tobacco: Age; Race; Gender; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Race; Gender; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Smokeless Tobacco: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A

Exhibit F.13 (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Stimulants, and Sedatives Snuff: Race; Gender; Census Region; MSA; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Alcohol, Marijuana, Cocaine, Hallucinogens, Pain Relievers	Chewing Tobacco: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Age*Gender; Gender*Race; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives Snuff: Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Census Region; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, and Stimulants	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Chewing Tobacco and Snuff 30-Day Frequency; Imputation-Revised Cigarette and Cigarette Daily at First Use

Exhibit F.14 Alcohol: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; Census Region; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Cigars, Smokeless Tobacco, and Pipes
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Alcohol 12-Month Frequency frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives**;** Intermediate Alcohol 12-Month Frequency
Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Alcohol 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco Age at First Use

Exhibit F.15 Alcohol: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Census Region; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Cigars, Smokeless Tobacco, and Pipes
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Alcohol 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Alcohol 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Alcohol 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco Age at First Use

Exhibit F.16 Alcohol: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Cigars, Smokeless Tobacco, and Pipes
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, and Pipes; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Alcohol Indicator
30-Day Frequency	Age; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Alcohol 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, and Pipes Recency; Lifetime Indicators of Marijuana, Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Alcohol 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Alcohol 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco Age at First Use

Exhibit F.17 Inhalants: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; Census Region; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, and Alcohol
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, and Marijuana; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Inhalant Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, and Marijuana; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Inhalant Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Inhalants 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Inhalants 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol and Cigars Age at First Use

Exhibit F.18 Inhalants: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; State Rank; MSA; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, and Alcohol
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Marijuana, and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Sex; Census Region; MSA; Imputation-Revised Cigarette, Cigar, Smokeless Tobacco, Pipes, Alcohol and Marijuana Recency; Lifetime Indicators of Pain Relievers, Tranquilizers, Hallucinogens, Stimulants, Sedatives, Cocaine, Crack and Heroin; Intermediate Past Month Inhalant Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, and Marijuana; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Inhalant Indicator
30-Day Frequency	Race; Imputation-Revised Cigarette, Cigar Recency; Lifetime Indicators of Pain Relievers, Tranquilizers, Stimulants	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Inhalants 12-Month Frequency
Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Inhalants 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol and Cigars Age at First Use

Exhibit F.19 Inhalants: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; State Rank; MSA; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, and Alcohol
Recency	Gender; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Gender; Imputation-Revised Cigarette, Marijuana and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes
12-Month Frequency	Age; Race; Census Region; MSA; Lifetime Indicators of Hallucinogens, Pain Reliever, Tranquilizers, and Stimulants; Intermediate Past Month Inhalant Indicator	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Race; Age*Gender; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars**,** Smokeless Tobacco, Pipe, Alcohol and Marijuana; Lifetime Indicators of Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalant Indicator
30-Day Frequency	Census Region; MSA; Lifetime Indicators of Pain Reliever; Intermediate Past Month Inhalant Indicator	Age; Gender; Race
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Inhalants 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol and Cigars Age at First Use

Exhibit F.20 Marijuana: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol and Inhalants
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives;	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes and Alcohol; Lifetime Indicators of , Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives;
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Marijuana 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Marijuana 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Marijuana 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants and Cigars Age at First Use

Exhibit F.21 Marijuana: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigar, Pipes, Alcohol and Inhalants
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco Pipes, and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives;
30-Day Frequency	Age; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Marijuana 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Marijuana 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Marijuana 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants and Cigars Age at First Use

Exhibit F.22 Marijuana: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigar, Pipes, Alcohol and Inhalants
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette and Alcohol Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Race; Gender; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Marital Status; Employment Status; Education Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco Pipes, and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives;
30-Day Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Hallucinogens, Inhalants, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Marijuana 12-Month Frequency
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Marijuana 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants and Cigars Age at First Use

Exhibit F.23 Hallucinogens: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants and Marijuana
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Hallucinogens Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Hallucinogens Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana, Inhalants and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana, Inhalants and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Hallucinogens 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Hallucinogens 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana and Cigars Age at First Use

Exhibit F.24 Hallucinogens: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants and Marijuana
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status;Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Hallucinogens Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Hallucinogens s Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Cigars, Smokeless Tobacco, Alcohol, Marijuana, Inhalants and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Cigars, Smokeless Tobacco, Alcohol, Marijuana, Inhalants and Pipes Recency; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Hallucinogens 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Hallucinogens 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana and Cigars Age at First Use

Exhibit F.25 Hallucinogens: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Census Region; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants and Marijuana
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Status; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Race; Gender; Age*Gender; Census Region; MSA; Education; Employment Status; Imputation-Revised Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Gender; Census Region; MSA; Lifetime Indicators of Cocaine, Heroin, Stimulants	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Hallucinogens Indicator
30-Day Frequency	Gender; Lifetime Indicators of Cocaine, Stimulants; Intermediate Hallucinogens 12-month Frequency	Age; Gender; Race; Gender*Age; Age*Race; Lifetime Indicators of Cocaine, Stimulants; Intermediate Hallucinogens 12-month Frequency
Age at First Use	Age Category; Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Hallucinogens 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana and Cigars Age at First Use

Exhibit F.26 Pain Relievers: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Census Region; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana and Hallucinogens
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3 Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, , Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Paine Relievers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens and Cigars Age at First Use

Exhibit F.27 Pain Relievers: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana and Hallucinogens
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2;Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, , Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator
30-Day Frequency	**N/A**	**N/A**
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Paine Relievers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens and Cigars Age at First Use

Exhibit F.28 Pain Relievers: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; MSA; State Rank; Census Region; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana and Hallucinogens
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, , Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Tranquilizers, Stimulants, and Sedatives; Intermediate Past Month Pain Relievers Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Paine Relievers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens and Cigars Age at First Use

Exhibit F.29 Tranquilizers: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census region; MSA; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens and Pain Relievers
Recency	Race; Gender; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Stimulants, and Sedatives; Intermediate Past Month Tranquilizers Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Stimulants, and Sedatives; Intermediate Past Month Tranquilizers Indicator
30-Day Frequency	**N/A**	**N/A**
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Tranquilizers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Cigars Age at First Use

Exhibit F.30 Tranquilizers: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens and Pain Relievers
Recency	Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Stimulants, and Sedatives; Intermediate Past Month Tranquilizers Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Stimulants, and Sedatives; Intermediate Past Month Tranquilizers Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Tranquilizers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Cigars Age at First Use

Exhibit F.31 Tranquilizers: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens and Pain Relievers
Recency	Race; Gender; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Race; Gender; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Gender; Census Region; MSA; Lifetime Indicators of Stimulants, Sedatives, Cocaine, and Crack; **Intermediate Past Month Tranquilizers Indicator**	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, Stimulants, and Sedatives; Intermediate Past Month Tranquilizers Indicator
30-Day Frequency	**N/A**	**N/A**
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Tranquilizers 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Cigars Age at First Use

Exhibit F.32 Stimulants: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Tranquilizers
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, and Sedatives; Intermediate Past Month Stimulants Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, and Sedatives; Intermediate Past Month Stimulants Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Stimulants 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Cigars Age at First Use

Exhibit F.33 Stimulants: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Tranquilizers
Recency	Age; Age^2; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Gender; Race; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Tranquilizers Lifetime Indicators of Cocaine, Crack, Heroin, and Sedatives; Intermediate Past Month Stimulants Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, and Sedatives; Intermediate Past Month Stimulants Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Stimulants 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Cigars Age at First Use

Exhibit F.34 Stimulants: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers and Tranquilizers
Recency	Age; Race; Gender; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Gender; Education; Employment Status; Census Region; MSA; Imputation-Revised Alcohol and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Gender; Census Region; MSA; Lifetime Indicators of Sedative; Intermediate Past Month Stimulants Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Employment Status; Education Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Alcohol; Lifetime Indicators of Cocaine, Crack, Heroin, and Sedatives; Intermediate Past Month Stimulants Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Age Category; Race; Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Stimulants 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Cigars Age at First Use

Exhibit F.35 Sedatives: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Census Region; MSA; State Rank; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Stimulants
Recency	Age; Race; Gender; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine	Age; Race; Gender; Census Region; MSA; State Rank; Imputation-Revised Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
12-Month Frequency	Gender; Race; MSA; Imputation-Revised Recency of Pain Relievers and Tranquilizers; Lifetime Indicators of Cocaine, Crack and Heroin; Intermediate Past Month Sedatives Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Alcohol; Lifetime Indicators of Cocaine, Crack and Heroin; Intermediate Past Month Sedatives Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Hallucinogens, Pain Relievers, Tranquilizers, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Sedatives 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Cigars Age at First Use

Exhibit F.36 Sedatives: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarette Lifetime Indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers and Stimulants
Recency	Age; Gender; Race; Age*Gender; Age*Race; Gender*Race; Education; MSA; State Rank; Imputation-Revised Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin	Age; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Hallucinogens, Stimulants Recency;	Age; Age^2; Age^3; Gender; Age*Gender; Marital Status; Education; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Alcohol; Lifetime Indicators of Cocaine, Crack, and Heroin; Intermediate Past Month Sedatives Indicator
30-Day Frequency	N/A	N/A
Age at First Use	Race; Census Region; MSA; Imputation-Revised Cigars, Smokeless Tobacco, Pipes, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Sedatives 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Cigars Age at First Use

Exhibit F.37 Sedatives: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Cigarettes Lifetime indicator; Intermediate Lifetime Indicator of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants
Recency	Gender; Imputation-Revised Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin	Age; Census Region; MSA; Education; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
12-Month Frequency	Gender	Age; Age^2; Age^3; Race; Gender; Age*Race; Gender*Race; State Rank; Gender*Age; Age*Race; Marital Status; Education; Employment Status; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, and Alcohol Recency
30-Day Frequency	**N/A**	**N/A**
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Alcohol, Marijuana, Cocaine, Crack, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Sedatives 12-Month Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Cigars' Age at First Use

Exhibit F.38 Cocaine: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Age*Gender; Race*Gender; Census Region; MSA; State Rank
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants, Heroin, and Crack
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives and Tranquilizers; Lifetime Indicator of Heroin, and Crack; Intermediate Past Month Cocaine Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Heroin, and Crack; Intermediate Past Month Cocaine Indicator
30-Day Frequency	Gender; Census Region; Imputation-Revised Recency of Cigars, and Alcohol; Lifetime Indicator of Crack;	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Crack, and Heroin; Intermediate Cocaine 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cocaine 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cigars' Age at First Use

Exhibit F.39 Cocaine: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Age*Gender; Race*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Education; Employment Status Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Imputation-Revised Recency of Cigarettes, Inhalants, Marijuana, Alcohol; Lifetime Indicator of Cigars, Smokeless Tobacco, Pipes, Heroin, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Heroin and Crack; Intermediate Past Month Cocaine Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Heroin and Crack; Intermediate Past Month Cocaine Indicator
30-Day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, and Alcohol; Lifetime Indicator of Heroin and Crack; Intermediate Cocaine 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Crack and Heroin; Intermediate Cocaine 12-Month Frequency
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cocaine 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives and Cigars Age at First Use

Exhibit F.40 Cocaine: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Age*Gender; Race*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants, and Heroin	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; State Rank; Imputation-Revised Cigarette, Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants, Crack, and Heroin
12-Month Frequency	Age Category; Gender; Census Region; MSA; Imputation-Revised Recency of Cigar, Alcohol, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Lifetime Indicator of Heroin; Intermediate Past Month Cocaine Indicator	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives and Tranquilizers; Lifetime Indicator of Heroin and Crack; Intermediate Past Month Cocaine Indicator
30-Day Frequency	Age Category; Gender; Intermediate Cocaine 12-Month Frequency	Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedatives, and Tranquilizers; Lifetime Indicator of Crack and Heroin; Intermediate Cocaine 12-Month Frequency
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cocaine 12-Month and 30-day frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Cigars Age at First Use

Exhibit F.41 Heroin: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Census Region; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Age*Gender; Race*Gender; MSA; State Rank;
Recency	Gender; Race; State Rank**;** Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Alcohol, and Marijuana; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack;
12-Month Frequency	Gender	Age; Age^2; Race; Gender; Gender*Race; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Cocaine, Crack; Intermediate Past Month Heroin Indicator
30-Day Frequency	Gender	Age; Race; Gender; Age*Gender; Age*Race; Census Region; MSA; State Rank
Age at First Use	Sex; Race; MSA; Imputation-Revised Smokeless Tobacco, Sedatives, Pain Relievers , and Heroin Recency	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol Age at First Use

Exhibit F.42 Heroin: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Census Region; Gender*Race; Age*Race; Age*Gender; Marital Status; Education; Employment Status; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine and Crack; Age*Gender; Race*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Race; Gender; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Alcohol, and Marijuana Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants , Cocaine	Age; Gender; Race; Census Region; MSA; State Rank; Employment Status; Imputation-Revised Cigarette Recency; Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Stimulants , Cocaine, and Crack
12-Month Frequency	Race; Gender; MSA; Imputation-Revised Alcohol, and Sedative Recency; Intermediate Past Month Heroin Indicator	Age; Age^2; Race; Gender; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; State Rank; Marital Status; Education; Employment Status; Imputation-Revised Recency of Cigarettes, Cigars, Smokeless Tobacco, Pipes, Inhalants, Marijuana, Alcohol, Hallucinogens, Pain Relievers, Stimulants, Sedative, Cocaine, Crack and Tranquilizers; Intermediate Past Month Heroin Indicator
30-Day Frequency	Race; Gender; Imputation-Revised Cigar Recency	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; State Rank; Education; Imputation-Revised Cigarettes
Age at First Use	Race; Census Region; Imputation-Revised Cigar, Pipes, Sedatives, and Heroin Recency	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Heroin 12-Month and 30-day frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack and Cigars Age at First Use

Exhibit F.43 Heroin: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Marital Status; Education; Employment Status Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Intermediate Lifetime Indicators of Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine and Crack; Age*Gender; Race*Gender; Census Region; MSA; State Rank; Marital Status; Education; Employment Status
Recency	Gender; Lifetime Indicators of Tranquilizers, Stimulants	Age; Age^2; Lifetime Indicators of Cigars, Hallucinogens, Stimulants ,Inhalants and Smokeless Tobacco
12-Month Frequency	Gender	Age; Age^2; Age^3; Gender; Race; Marital Status; Education; Employment Status; State Rank; Gender*Race; Gender*Age; Age*Race
30-Day Frequency	Age; Race; Gender; Census Region; MSA	Age; Gender
Age at First Use	Age; Race; Gender; Census Region; MSA; Imputation-Revised Cigarette, Marijuana, and Cocaine Recency	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; Age^2*Race; Age^2*Gender; MSA; Marital Status; Education; Employment Status; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Heroin 12-Month and 30-day frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack and Cigars Age at First Use

F.4 Health Insurance Variables

Exhibit F.44 Health Insurance, 2000 Method: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.45 Health Insurance, 2000 Method: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.46 Health Insurance, 2000 Method: 26 to 64 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.47 Health Insurance, 2000 Method: 65+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Gender; Race; Marital Status	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.48 Health Insurance, 1999 Method: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.49 Health Insurance, 1999 Method: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.50 Health Insurance, 1999 Method: 26 to 64 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

Exhibit F.51 Health Insurance, 1999 Method: 65+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Overall Health Insurance	Age; Gender; Race; Marital Status	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size
Private Health Insurance	N/A	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic Population; Percent Non-Hispanic Black Population; Percent of Owner-Occupied Households; Household Size

F.5 Income Variables

Exhibit F.52 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12 to 17	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent Owner Occupied in Segment; Marital Status; Education; 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	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent Owner Occupied in Segment; Marital Status; Education; 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+	Age; Age^2; Gender; Race; Gender*Race; Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent Owner Occupied in Segment; Marital Status; Education; 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

Exhibit F.53 Dichotomous Income Indicators in Predictive Mean Modeling: 12 to 17 Year Olds

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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
Supplemental Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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

Exhibit F.53 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
# Welfare Months	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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
Total Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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

Exhibit F.54 Dichotomous Income Indicators in Predictive Mean Modeling: 18 to 25 Year Olds, 26 to 64 Year Olds, and 65+ Year Olds (Covariates Are the Same)

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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

Exhibit F.54 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
# Welfare Months	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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; Intermediate Family Food Stamps
Investment Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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; Intermediate Family Food Stamps

Exhibit 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^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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 and 26 to 64	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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; 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+	Age; Age^2; Gender; Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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)

Exhibit 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^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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)
All others	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent 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 Household Composition Variables

Exhibit F.57 Household Composition: 12 to 17 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total people in household (Screener)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener)
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size
Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in household	Age; Age^2; Gender; Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in household
Other family present in Household (IRFAMSKP)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Exhibit F.58 Household Composition: 18 to 25 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size
Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education Level
Other family present in Household (IRFAMSKP)	Age; Age^2; Gender; Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level

Exhibit F.59 Household Composition: 26 to 64 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size
Imputation-Revised Number of Persons Greater Than 64 in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education Level
Other Family Present in Household (IRFAMSKP)	Age; Age^2; Gender; Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level

Exhibit F.60 Household Composition: 65+ Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Gender; Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education Level
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Gender; Race; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised Household Size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size
Imputation-Revised Number of Persons Greater Than 64 Years old in Household (IRHH65)	Age; Gender; Race; Imputation-Revised Household Size; Marital Status	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education Level
Other Family Present in Household (IRFAMSKP)	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education Level

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 variables for which imputations were implemented, whether the predictive mean neighborhood (PMN) 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 according to logical and likeness constraints, where the likeness constraints were loosened if donors could not be found, but logical constraints could not be loosened. Because logical constraints (summarized in **Appendix H**) cannot be loosened, the attempt to find a donor under those constraints is either successful or not successful; there is no opportunity to loosen the constraints.[80] Such an opportunity does exist, however, with likeness constraints. If no donors were available under the most stringent set of constraints, constraints were loosened, one at a time, until a donor could be 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.

Although statistical imputation of the drug use or income variables could not proceed 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 exhibits that follow, this variable is identified as the "State rank" for the drug use and income variables. It was used as a likeness constraint, where the set of eligible donors for each recipient was restricted so that donors and recipients were both from States with the same State rank.

The phrase "Donor's predicted means each within *x* percent of recipient's predicted means" appears in each of the exhibits corresponding to a multivariate imputation, and the phrase "Donor's predicted mean within *x* percent of recipient's predicted mean" appears in each

[80] Logical constraints define what is normally referred to as an "imputation class."

of the univariate imputation exhibits. In either case, it represents one of the likeness constraints. It also defines the neighborhood. Once this constraint is loosened, the neighborhood is abandoned and the candidate with the predicted mean closest to the recipient's, subject to the constraints that are still on the pool of donors, is chosen as the donor.

G.2 Demographics

G.2.1 Race Variables

Exhibit G.1 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	318	69	33
(A) Donor's predicted means within 5 percent of recipient's predicted means	943	727	513
None	188	719	667

G.2.2 Hispanic Origin Variables

Exhibit G.2 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	22	3	1
(A) Donor's predicted mean within 5 percent of recipient's predicted mean	13	4	5

G.2.3 Marital Status Variables

Exhibit G.3 Marital Status Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean	4	5	14
None	3	1	3

G.2.4 Hispanic Group Variables

Exhibit G.4 Hispanic Group 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	4	9	1
(A) Donor's predicted means within 5 percent of recipient's predicted means	27	14	8
None	9	9	5

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. The exhibits 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

Exhibit G.5 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	427	88	104
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means with matches for multiple cases delta	34	41	53
(A) State rank of donor = State rank of recipient	44	34	64

G.3.2 Likeness Constraints for Recency and Frequency Imputation, by Drug

Exhibits G.6 to **G.18** present information on the likeness constraints for recency and frequency imputation for the following drugs: tobacco (i.e., cigarettes, cigars, and smokeless tobacco [chewing tobacco and snuff]), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, sedatives, and stimulants), cocaine, and heroin.

Exhibit G.6 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	534	167	36
(A) State rank of donor = State rank of recipient	35	17	31

Exhibit G.7 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	339	229	66
(A) State rank of donor = State rank of recipient	29	21	17

Exhibit G.8 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	150	88	8
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	14	7	1
(A) State rank of donor = State rank of recipient	61	33	18

Exhibit G.9 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	708	762	694
(A) State rank of donor = State rank of recipient	327	176	179

Exhibit G.10 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	45	9	0
(A) State rank of donor = State rank of recipient	210	49	14

Exhibit G.11 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	126	95	26
(A) State rank of donor = State rank of recipient	196	152	77

Exhibit G.12 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 and PCP agree with recipient's recencies (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means	37	37	1
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	11	10	1
(A) State rank of donor = State rank of recipient	163	136	44

Exhibit G.13 Analgesics 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	99	50	13
(A) State rank of donor = State rank of recipient	137	91	67

Exhibit G.14 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	15	10	1
(A) State rank of donor = State rank of recipient	46	26	29

Exhibit G.15 Sedatives 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	2	2	1
(A) State rank of donor = State rank of recipient	16	14	8

Exhibit G.16 Stimulants 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 methamphetamines agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5 percent of recipient's predicted means	49	55	9
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	20	36	2
(A) State rank of donor = State rank of recipient	108	68	52

Exhibit G.17 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	4	9	0
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	0	1	0
(A) State rank of donor = State rank of recipient	62	79	35
None	2	0	0

Exhibit G.18 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	0	0	0
(A) State rank of donor = State rank of recipient	11	14	4

G.3.3 Likeness Constraints for Age at First Use Imputation, by Drug

Exhibits G.19 to G.31 present information on the likeness constraints for age at first use (AFU) imputation for the following drugs: tobacco (i.e., cigarettes, cigars, and smokeless tobacco [chewing tobacco and snuff]), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, sedatives, and stimulants), cocaine, and heroin.

Exhibit G.19 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	565	239	241
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	1	0	15
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	2	0	8
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient*	0	0	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.20 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	331	273	261
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	4	0	26
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	1	1	30
(A) Age of donor = Age of recipient	0	0	3

Exhibit G.21 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 did not use in the past year, donor must not have used in the past year (these checks are only done for the drugs for which the recipient has missing AFU) (D) Donor's predicted mean within 5 percent of recipient's predicted mean	161	137	57
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (these checks are only done for the drugs for which the recipient has missing AFU) (C) Donor's predicted mean within 5 percent of recipient's predicted mean	43	9	24
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (these checks are only done for the drugs for which the recipient has missing AFU)	19	3	30
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	5
(A) AFU of donor ≤ Age of recipient *	0	0	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.22 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	548	278	408
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	23
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	1	0	13
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.23 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	250	60	30
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	0	3
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	2	2	5
(A) Age of donor = Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient*	0	0	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.24 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	138	122	105
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	5
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	0	1	7
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,* Age of donor ≥ Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient	0	0	1

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.25 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 did not use in the past year, donor must not have used in the past year (this check is done for overall hallucinogens, LSD, and PCP) (D) Donor agrees with recipient with respect to lifetime use for both LSD and PCP (E) Donor's predicted mean within 5 percent of recipient's predicted mean	106	81	44
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (this check is done for overall hallucinogens, LSD, and PCP) (C) Donor agrees with recipient with respect to lifetime use for both LSD and PCP (D) Donor's predicted mean within 5 percent of recipient's predicted mean	9	4	10
(A) Age of donor = Age of recipient (B) Donor's predicted mean within 5 percent of recipient's predicted mean	3	3	6
(A) Age of donor = Age of recipient	4	5	9
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens),* Age of donor ≥ Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens)*	0	0	6
(A) AFU of donor ≤ Age of recipient, the absolute difference between the donor's age and the recipient's age ≤ 20 years	0	0	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.26 Analgesics 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	279	176	111
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	10	1	28
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	3	0	27
(A) Age of donor = Age of recipient	0	0	4
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	10
(A) AFU of donor ≤ Age of recipient	0	0	2

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.27 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	56	40	37
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	2	2	5
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	1	1	15
(A) Age of donor = Age of recipient	0	0	2
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.28 Sedatives 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 used in the past year, donor must have, too; if recipient did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	17	16	19
(A) Age of donor = Age of recipient (B) If recipient used in the past year, donor must have, too; if recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	6	1	7
(A) Age of donor = Age of recipient (B) If recipient used in the past year, donor must have, too; if recipient did not use in the past year, donor must not have used in the past year	5	5	7
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient	0	0	2

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.29 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 did not use in the past year, donor must not have used in the past year (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	110	64	34
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (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	7	2	12
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (this check is done for both overall stimulants and methamphetamines) (C) Donor agrees with recipient with respect to lifetime use for methamphetamines (checked only if recipient is a nonrespondent for methamphetamines AFU) (D) Donor's predicted mean within 5 percent of recipient's predicted mean	2	0	3
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (this check is done for both overall stimulants and methamphetamines) (C) Donor agrees with recipient with respect to lifetime use for methamphetamines (checked only if recipient is a nonrespondent for methamphetamines AFU)	7	3	10
(A) Age of donor = Age of recipient (B) Donor agrees with recipient with respect to lifetime use for methamphetamines (checked only if recipient is a nonrespondent for methamphetamines AFU)	0	1	0
(A) Donor is at least as old as recipient, but no more than 20 years older than recipient (B) AFU of donor ≤ Age of recipient (for overall stimulants)*	0	0	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.30 Cocaine 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 did not use in the past year, donor must not have used in the past year (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	25	41	32
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (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	1	2	5
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (this check is done for both overall cocaine and crack) (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU) (D) Donor's predicted mean within 5 percent of recipient's predicted mean	0	0	1
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (this check is done for both overall cocaine and crack) (C) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)	6	0	7
(A) Age of donor = Age of recipient (B) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)	0	0	2
(A) Donor is at least as old as recipient, but no more than 20 years older than recipient (B) AFU of donor ≤ age of recipient (for overall stimulants)*	1	0	2

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit G.31 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 did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5 percent of recipient's predicted mean	3	6	1
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year (C) Donor's predicted mean within 5 percent of recipient's predicted mean	4	0	0
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	4	0	4
(A) Age of donor = Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient*	0	0	0

* Although this is a logical constraint, it is included for the sake of clarity.

G.4 Health Insurance Variables

Exhibits G.32 and G.33 present information on the likeness constraints for the health insurance variables. There are two tables because there were eight MPMN programs: one for each age group using the "2000 method," and one for each age group using the "1999 method." See **Chapter 7** for an explanation of the two methods. Although the likeness constraints were the same in all MPMN programs, the lists of item nonrespondents were slightly different for the two methods.

Exhibit G.32 2000 Method, Health Insurance (IRINSUR2) and Private Health Insurance (IRPINSUR) Imputations

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	651	237	68	11
(A) Age of donor = Age of recipient	9	6	24	3

Exhibit G.33 1999 Method, Health Insurance (IRINSUR) and Private Health Insurance (IRPINSUR) Imputations

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	655	228	64	7
(A) Age of donor = Age of recipient	6	5	28	7

G.5 Binary Variable Phase

G.5.1 Income Variables

The item nonrespondents for the binary income variables were divided into one of two classes. Those with missing values for any of the welfare-correlated variables (family food stamps, personal/other family welfare payments, personal/other family welfare services, personal/other family interest, personal/other family total income, and family months-on-welfare) went through the usual MPMN process, with likeness constraints and so on. All other item nonrespondents, along with those in the first category for whom a donor could not be found, were assigned provisionally imputed values for all missing variables.

Exhibit G.34 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) (C) If recipient is missing only one edited variable of a (personal, other-family) pair, donor's value is equal to the recipient's value for the nonmissing one (D) Donor's predicted means within 5 percent of recipient's predicted means for all missing family variables	2067	1399	814	179
(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) (C) If recipient is missing only one edited variable of a (personal, other-family) pair, donor's value is equal to the recipient's value for the nonmissing one	913	664	633	275
(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) (C) If recipient is missing only one edited variable of a (personal, other-family) pair, donor's value is equal to the recipient's value for the nonmissing one	26	27	24	10
(A) Donor's values for welfare-correlated edited binary income variables are the same as recipient's values (when nonmissing) (B) If recipient is missing only one edited variable of a (personal, other-family) pair, donor's value is equal to the recipient's value for the nonmissing one	0	2	5	2
(A) If recipient is missing only one edited variable of a (personal, other-family) pair, donor's value is equal to the recipient's value for the nonmissing one	2	0	1	1
Use provisionally imputed values for whichever income variables are missing	195	177	62	11

G.5.2 Specific Category Phase

Exhibit G.35 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	4,961	3,754	2,627	761
(A) Donor's predicted mean within 10 percent of recipient's predicted mean (B) FINC2 of donor ≥ PINC2 of recipient, if not missing* (C) PINC2 of donor ≤ FINC2 of recipient, if not missing*	11	2	3	5

* Although this is a logical constraint, it is included for the sake of clarity.

G.6 Household Composition (Roster) Variables

Exhibit G.36 IRHHSIZE Imputations

Likeness Constraints	Frequency			
	12-17	**18-25**	**26-64**	**65+**
Donor's predicted mean within 5 percent of recipient's predicted mean	67	46	54	5

Exhibit G.37 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	99	84	95	10
(A) IRHHSIZE of donor = IRHHSIZE of recipient	0	0	1	0

Exhibit G.38 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	121	94	98	10
(A) IRHHSIZE of donor = IRHHSIZE of recipient	0	0	1	0

Exhibit G.39 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	86	103	116	7
(A) IRKID17 of donor = IRKID17 of recipient	0	0	0	1

Appendix H: Missingness Patterns

Appendix H: Missingness Patterns

H.1 Introduction

The predictive mean neighborhood (PMN) imputation method was applied to many variables in the 2000 National Household Survey on Drug Abuse (NHSDA). 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). **Exhibit H.1** gives examples of variables, which were imputed using each of the four methods.

Exhibit H.1 Lists of Variables Imputed Using Each of the Four Methods of PMN

	Variables Imputed One at a Time (Univariate Assignment)	Variables Imputed in Set (Multivariate Assignment)
Predictive mean vector has one element (univariate matching)	IRHOIND, IRHHSIZE, IRHH65, IRKID17, IRFAMSKP, IR*xxx*AGE	{IRPINC2, IRFINC2, IRFAMIN2}
Predictive mean vector has more than one element (multivariate matching)	IRMARIT, IRHOGRP3, EMPSTAT3, EMPSTATY, IRNWRACE	{IR*xxx*RC, IR*xxx*FY, IR*xxx*FM}, {lifetime drug use}, {IRINSUR, IRPINSUR}, {IRINSUR2, IRPINSUR}, {binary source of income}

Note: The xxx refers to the three-letter abbreviation for each drug in turn (e.g., CIG for cigarettes).

For many of these variables, the item nonrespondents were segregated into missingness patterns, which are simply patterns of nonresponse. Missingness patterns were created in two ways. The first was applied to variables that underwent multivariate assignment: They can be segregated into missingness patterns based on which variables were missing. The second way occur when logical editing can restrict an item nonrespondent to only a subset of the variable's possible values. For example, logical editing can sometimes restrict a lifetime user of a drug to past year use; in these cases, it is known that the recipient should receive a final imputed value of 1 or 2 for drug recency. This can happen for any variables undergoing 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 differ between missingness patterns. The exhibits in this appendix specify, for each missingness pattern, the number of item nonrespondents exhibiting the pattern, the set of logical constraints applied to the potential donors, and the elements of the predictive mean vector used to calculate the Mahalanobis distance from recipient to potential donor.

Often, differences among missingness patterns with respect to the predictive mean vector are due to the use of conditional probabilities. If something about the item nonrespondent is known, probabilities conditioned on what is known are used. For example, only past month users are included in models for 30-day frequency. Therefore, the predictive means calculated using these models are conditional on past month use of the drug. If an item nonrespondent is missing both recency and 30-day frequency for that drug, probabilities conditional on lifetime use, not on past month use, are used for the predictive mean vector. Conditional probabilities often result if the variables imputed using a multivariate assignment method are related in a hierarchical manner, such as overall health insurance and private health insurance, or if partial information is available about an item nonrespondent, such as the cases where it is known that the recipient is a past year user of a drug, but it is unknown whether he or she is a past month user.

In 2001, the use of conditional probabilities will be extended to both the health insurance variables and the source of income variables. In the case of the health insurance variables, whenever overall insurance is missing and private health insurance is not, the item nonrespondent not having private health insurance is conditional: If he or she had private health insurance, he or she would necessarily have overall health insurance. Conversely, whenever overall insurance is nonmissing and private insurance is missing, it is known that the respondent must have overall insurance, which can serve as the conditional item.

In the case of the source of income variables, there is a hierarchical relationship among the welfare payments, welfare services, and months-on-welfare variables. The model for months-on-welfare includes only welfare recipients (welfare payments, welfare services, or both), so the probabilities estimated by the model are conditional on the receipt of welfare. For item nonrespondents missing all three variables, unconditional probabilities for months-on-welfare will be calculated in 2001; for item nonrespondents with other missingness patterns, different conditional probabilities can be calculated.

In 2002, the use of conditional probabilities will be extended to the employment status variable, EMPSTATY, for item nonrespondents who can be restricted to either full-time employment or part-time employment.

Section H.2 shows the logical constraints associated with each missingness pattern for each variable/set of variables that used missingness patterns. It also reports the number of item nonrespondents exhibiting each missingness patterns. **Section H.2.1** deals with drug lifetime use,

Section H.2.2 deals with drug recency and frequency, **Section H.2.3** deals with the health insurance variables, and **Section H.2.4** deals with the source of income variables.

Section H.3 shows the elements of the predictive mean vector that were used in each missingness pattern. **Section H.3.1** deals with drug lifetime use, **Section H.3.2** deals with drug recency and frequency, **Section H.3.3** deals with the health insurance variables, and **Section H.3.4** deals with the source of income variables.

H.2 Exhibits Showing Missingness Patterns and the Restrictions on the Set of Potential Donors

A few items to note regarding the exhibits in **Section H.2** are as follows.[81] In the missingness pattern section, no entry in the columns indicates that all information is available; an entry of "Missing" indicates that all information is missing. Other entries in the missingness pattern section give the available information, indicating that the information is partially missing. However, if the entry is in parentheses, all information is present and was thought to be useful for the reader.

H.2.1 Drug Lifetime Use

There are a large number of missingness patterns for drug lifetime use. The response to the gate question for cigarettes must be nonmissing for the survey to be considered complete, but any combination of the other lifetime drug variables may be missing. There are 14 other gate questions in the questionnaire, plus several subgate questions.

There are no logical constraints for any of these missingness patterns.

H.2.2 Drug Recency and Frequency

See **Exhibits H.2 to H.19** on the following pages. Please note that in this section, pain relievers, sedatives, and tranquilizers have identical missingness patterns and are therefore presented in the same exhibit.

[81] Many exhibits 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."

Exhibit H.2 Constraints for Tobacco (Cigarettes and Cigars)

Constraint #	Logical Constraint
Tob1	If the difference between the recipient's current age and his/her age at first use is 2 years or less, the recipient must have used within the past 3 years (a recency category of 1, 2, or 3)
Tob2	Recipient cannot be a past month user (recency cannot equal 1)
Tob3	Recipient must used drug within the past year (recency = 1 or 2)
Tob4	Recipient must be a past month user (recency = 1)
Tob5	If the recipient was never a daily user of cigarettes (CG15=2), the donor's 30-day cigarette frequency cannot equal 30
Tob6	If recipient's age at first use equals his/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/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/her birthday (inclusive)

Exhibit H.3 Cigarette User Restrictions

Missingness Pattern			Number of Cases	Logical Constraints
#	Recency	30-Day Frequency		
1	Past year	Missing	15	(Tob1), (Tob5)
2	Missing (lifetime use imputed)	Missing	0	(Tob1), (Tob5)
2	Missing (lifetime use known)	Missing	57	
3	(Past month)	Missing	61	(Tob1), (Tob4), (Tob5), (Tob6)
4	Not past year		363	(Tob1), (Tob3), (Tob5)
5	Not past month		324	(Tob1), (Tob2), (Tob5)
6	30-day frequency logically assigned based on estimated value, no missing values.		0	(Tob1), (Tob5)
	Lifetime user, nothing missing		40,076	(None)
	Imputed to lifetime nonuse		0	(None)
	Lifetime nonuser, nothing missing		30,868	(None)

Exhibit H.4 Cigar User Restrictions

Missingness Pattern			Number of Cases	Logical Constraints
#	Recency	30-Day Frequency		
1	Past year	Missing	17	(Tob1)
2	Missing (Lifetime use imputed)	Missing	5	(Tob1)
2	Missing (Lifetime use known)	Missing	37	
3	(Past month)	Missing	20	(Tob1), (Tob4), (Tob6)
4	Not past year		258	(Tob1), (Tob3)
5	Not past month		364	(Tob1), (Tob2)
6	30-day frequency logically assigned based on estimated value, no missing values.		63	(Tob1)
	Lifetime user, nothing missing		21,480	
	Imputed to lifetime nonuse		7	
	Lifetime nonuser, nothing missing		49,513	

Exhibit H.5 Constraints for Smokeless Tobacco (Chewing Tobacco and Snuff)

Constraint #	Description
SLT1	If the difference between the recipient's current age and his/her age at first chew use is 2 years or less, the recipient must have used chew within the past 3 years (a recency category of 1, 2, or 3)
SLT2	If the difference between the recipient's current age and his/her age at first snuff use is 2 years or less, the recipient must have used snuff within the past 3 years (a recency category of 1, 2, or 3)
SLT3	Donor's not a chew user, then recipient must also not be a chew user (and vice versa)
SLT4	Donor's 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/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/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/her birthday (inclusive)
SLT6	If recipient's age at first snuff use equals his/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/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/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 equal to 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)

Exhibit H.6 Smokeless Tobacco Users (Snuff and Chewing Tobacco) Restrictions

Missingness Pattern						
#	**Chew Recency**	**Snuff Recency**	**Chew 30-Day Freq.**	**Snuff 30-Day Freq.**	**Number of Cases**	**Logical Constraints**
1	(Past month)	(Past month)	Missing	Missing	2	(SLT1-SLT4), (SLT5-SLT8)
2	(Past month)		Missing		4	(SLT1-SLT4), (SLT5), (SLT7), (SLT9)
3		(Past month)		Missing[1]	4	(SLT1-SLT4), (SLT6), (SLT8), (SLT10)
4		Missing (Lifetime use known)		Missing	2	(SLT1-SLT4), (SLT6), (SLT10)
4		Missing (Lifetime use imputed)		Missing	3	(SLT1-SLT4), (SLT6), (SLT10)
5	(Past month)	Missing (Lifetime use known)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT10)
5	(Past month)	Missing (Lifetime use imputed)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT10)
6	Missing (lifetime use known)		Missing		11	(SLT1-SLT4), (SLT5), (SLT9)
6	Missing (lifetime use imputed)		Missing		2	(SLT1-SLT4), (SLT5), (SLT9)
7	Missing (lifetime use known)	(Past month)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT8)
7	Missing (lifetime use imputed)	(Past month)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT8)
8		Past year		Missing	2	(SLT1-SLT4), (SLT10-SLT11)
9	Past year		Missing		9	(SLT1-SLT4), (SLT5), (SLT8), (SLT12)
10	Missing (lifetime use known)	Missing (Lifetime use known)	Missing	Missing	2	(SLT1-SLT4), (SLT5-SLT6)

Exhibit H.6 (continued)

Missingness Pattern						
#	**Chew Recency**	**Snuff Recency**	**Chew 30-Day Freq.**	**Snuff 30-Day Freq.**	**Number of Cases**	**Logical Constraints**
10	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	1	(SLT1-SLT4), (SLT5-SLT6)
10	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6)
10	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6)
11	Not past year				84	(SLT1-SLT4), (SLT8), (SLT13)
12		Not past year			45	(SLT1-SLT4), (SLT10), (SLT14)
13	Not past year	Not past year			9	(SLT1-SLT4), (SLT13-SLT14)
14	Not past month				138	(SLT1-SLT4), (SLT9), (SLT15)
15		Not past month			46	(SLT1-SLT4), (SLT10), (SLT16)
16	Not past month	Not past month			15	(SLT1-SLT4), (SLT15-SLT16)
17	Not past month	(Past month)		Missing	0	(SLT1-SLT4), (SLT6), (SLT8), (SLT15)
18	(Past month)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT7), (SLT16)
19	Not past month	Missing (lifetime use known)		Missing	0	(SLT1-SLT4), (SLT6), (SLT15)
19	Not past month	Missing (lifetime use imputed)		Missing	0	(SLT1-SLT4), (SLT6), (SLT15)
20	Missing (lifetime use known)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT16)
20	Missing (lifetime use imputed)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT16)
21	Not past month	Not past year			0	(SLT1-SLT4), (SLT14-SLT15)
22	Not past year	Not past month			1	(SLT1-SLT4), (SLT13), (SLT16)

Exhibit H.6 (continued)

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints
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	(SLT1-SLT4)
					0	
	Lifetime user, nothing missing				12,448	
	Imputed to lifetime nonuse				25	
	Lifetime nonuser, nothing missing				58,911	

Exhibit H.7 Pipe User Restrictions

Missingness Pattern		Number of Cases	Constraints
#	Recency		
1	Missing (lifetime use imputed)	2	(None)
1	Missing (lifetime use known)	6	(None)
	Lifetime user, nothing missing	6,420	
	Imputed to lifetime nonuse	10	
	Lifetime nonuser, nothing missing	65,326	

Exhibit H.6 (continued)

Exhibit H.8 Constraints for Various Drugs

Drug	Constraint #	Constraint
Alc, Mrj, Inh, Anl, 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) 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 or equal to than 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 (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 (30)
Alc, Mrj, Inh, Anl, Trn, Sed	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) 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 freq (2) if the recipient is not a past month user but a past year user, it must be at least 1
Alc, Mrj, Inh, Anl, Trn, Sed	C3	(Recipient's proportion of past year use * max number of days could have used in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Alc, Mrj, Inh	C4	(Donor's proportion of past year use * recipient's number of days could have used in past year) greater than or equal to 30-day use
Alc, Mrj, Inh	C5	Donor's 30-day use less than number of days between recipient's interview date and birthday (+1)
Alc, Mrj, Inh	C6	Donor's 30-day use less than the recipient's maximum number of days could have used in past 30 days
Alc, Mrj, Inh	C7	Donor's 30-day use greater than the recipient's minimum number of days could have used in past 30 days
Alc, Mrj, Inh	C8	Donor's 30-day use greater than recipient's DR5DAY (# days had 5+ drinks in past 30 days)
Alc, Mrj, Inh	C9	Donor's 30-day use greater than (donor's proportion of past year use * recipient's max number of days could have used in past year [335])
Alc, Mrj, Inh, Anl, Trn. Sed	C10	Donor must be a past month user (recency = 1)
Alc, Mrj, Inh	C11	If recipient's age at first use equals his/her current age, the donor's 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first drug use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)

Exhibit H.8 (continued)

Drug	Constraint #	Constraint
Alc, Mrj, Inh	C12	If recipient's age at first use equals his/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/her interview date and date of first drug use (+1) 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/her interview date and birthday (+1)
Alc, Mrj, Inh	C13	Recipient's estimated 30-day frequency is not given/legitimately skipped (estimated frequency not equal to 1-6)
Alc, Mrj, Inh	C14	If recipient's age at first use equals his/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/her interview date and date of first drug use (-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 (-29)
Alc, Mrj, Inh, Anl, Trn. Sed	C15	Donor must be a past year (but not past month) user (recency = 2)
Alc, Mrj, Inh	C16	Donor's DR5DAY values is less than recipient's 30-day frequency
Alc, Mrj, Inh	C17	If recipient's age at first use equals his/her current age, (1) donor's DR5DAY must be less than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's DR5DAY must be less than recipient's days between his/her interview date and birthday (+1)
Alc, Mrj, Inh, Anl, Trn. Sed	C18	Donor must be a past month or past year (but not past month) use (recency = 1 or 2)
Alc, Mrj, Inh	C19	Donor's proportion of past year use * recipient's max number of days could have used in past year greater than donor's 30-day frequency
Alc, Mrj, Inh, Her	C20	If recipient's age at first use equals his/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/her interview date and date of first drug use (-365) 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/her interview date and birthday (-365)
Alc, Mrj, Inh, Her	C21	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 max number of days could have used in past year (30 + 30-day frequency)

Exhibit H.9 Alcohol User Restrictions

Missingness Pattern					Number of Cases	Logical Constraints
#	Recency	12-Month Freq.	30-Day Freq.	30-Day Binge Drink		
1	(Past month)	Missing	Missing		20	(C1-C13)
2	(Past month)		Missing		275	(C5-C8), (C10), (C11), C13
3	(Past month)	Missing			170	(C1-C4), (C10), (C12)
4	(Past year but not past month)	Missing			220	(C1-C3), (C14), (C15)
5	(Past month)			Missing	938	(C10), (C16), (C17)
6	(Past month)		Missing	Missing[1]	11	(C5-C7), (C10), (C11), (C13)
7	(Past month)	Missing		Missing	104	(C1-C4), (C10), (C12), (C16), (C17)
8	(Past month)	Missing	Missing	Missing	46	(C1-C4), (C5-C7), (C9-C13)
9	Past Year		Missing	Missing	485	(C5-C7), (C11), (C13, C15)
10	Past year	Missing	Missing	Missing	62	(C1-C3), (C5-C9), (C11-C14), (C18)
11	Lifetime (imputed)	Missing	Missing	Missing	23	(C1-C7), (C9), (C11-C14) (C1-C70, (C9), (C11-C14)
11	Lifetime (known)	Missing	Missing	Missing	492	
	Lifetime user, nothing missing				47,426	
	Imputed to lifetime nonuse				8	
	Lifetime nonuser, nothing missing				21,484	

.

Exhibit H.10 Marijuana User Restrictions

Missingness Pattern				Number of Cases	Constraints
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	20	(C1-C7), (C9-C13)
2	(Past month)		Missing	8	(C5-C7), (C10), (C11), (C13)
3	(Past month)	Missing		61	(C1-C4), (C10), (C12)
4	(Past year but not past month)	Missing		68	(C1-C3), (C13), (C14)
5	Past year		Missing	94	(C5-C7), (C11), (C13), (C18)
6	Past year	Missing	Missing	97	(C1-C3), (C5-C7), (C9), (C11-C14), (C18), (C19)
7	Missing (lifetime use imputed)	Missing	Missing	64	(C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20) (C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20)
7	Missing (lifetime use known)	Missing	Missing	260	
	Lifetime user, nothing missing			23,305	
	Imputed to lifetime nonuse			60	
	Lifetime nonuser, nothing missing			47,727	

Exhibit H.11 Inhalants User Restrictions

Missingness Pattern				Number of Cases	Constraints
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	4	(C1-C7), (C10), (13)
2	(Past month)		Missing	4	(C6-C8), (C10), (C13)
3	(Past month)	Missing		6	(C1-C4), (C10)
4	(Past year not past month)	Missing		23	(C1-C3), (C18)
5	Past year		Missing	28	(C5-C7), (C9),(C13), (C18)
6	Past year	Missing	Missing	4	(C1-C3), (C5-C7), (C9), (C13), (C18)
7	Missing (lifetime use imputed)	Missing	Missing	5	(C1-C3), (C5-C7), (C9), (C13)
7	Missing (lifetime use known)	Missing	Missing	252	(C1-C3), (C5-C7), (C9), (C13)
	Lifetime user, nothing missing			6,580	
	Imputed to lifetime nonuse			119	
	Lifetime nonuser, nothing missing			64,739	

Exhibit H.12 Heroin User Restrictions

Missingness Pattern				Number of Cases	Constraints
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	0	(C1-C7), (C9), (C10-C13), (C21)
2	(Past month)		Missing	0	(C5-C7), (C10), (C13)
3	(Past month)	Missing		0	(C1-C4), (C10), (C21)
4	(Past year but not past month)	Missing		0	(C1-C3), (C15)
5	Past year		Missing	7	(C5-C7), (C9), (C13), (C18)
6	Past year	Missing	Missing	4	(C1-C3), (C5-C7), (C9), (C13), (C18), (C21)
7	Missing (lifetime use imputed)	Missing	Missing	1	(C1-C3), (C5-C7), (C9), (C13), (C21)
7	Missing (lifetime use known)	Missing	Missing	17	(C1-C3), (C5-C7), (C9), (C13), (C21)
	Lifetime user, nothing missing			642	
	Imputed to lifetime nonuse			40	
	Lifetime nonuser, nothing missing			71,053	

Exhibit H.13 Users of Pain Relievers, Tranquilizers, and Sedatives

Missingness Pattern				
#	Recency	12-Month Frequency	Number of Cases	Constraints
1	(Past month)	Missing	Pain relievers: 47	(C1-C3), (C10)
			Tranquilizers: 7	
			Sedatives: 1	
2	(Past year but not past month)	Missing	Pain relievers: 47	(C1-C3), (C15)
			Tranquilizers: 18	
			Sedatives: 4	
3	Past year		Pain relievers: 3	(C18)
			Tranquilizers: 1	
			Sedatives: 1	
4	Past year	Missing	Pain relievers: 15	(C1-C3), (C18)
			Tranquilizers: 5	
			Sedatives: 1	
5	Missing (lifetime use imputed)	Missing	Pain relievers: 21	(C1-C3), (C18) (C1-C3), (C18)
			Tranquilizers: 6	
			Sedatives: 9	
5	Missing (lifetime use known)	Missing	Pain relievers: 324	
			Tranquilizers: 90	
			Sedatives: 27	
	Lifetime user, nothing missing		Pain relievers: 6,826	
			Tranquilizers: 3,489	
			Sedatives: 1,257	
	Imputed to lifetime nonuse		Pain relievers: 259	
			Tranquilizers: 176	
			Sedatives: 188	
	Lifetime nonuser, nothing missing		Pain relievers: 64,222	
			Tranquilizers: 67,972	
			Sedatives: 70,276	

Exhibit H.14 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) 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 cocaine month user, the past year cocaine frequency must be less than or equal to the maximum period (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) 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 freq (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) less than or equal to the number of days between recipient's interview date and birthday (+1)
Coc5	(Donor's proportion of past year cocaine use * recipient's number of days could have used cocaine in past year) greater than or equal to 30-day use
Coc6	Donor's 30-day cocaine use less than number of days between recipient's interview date and birthday (+1)
Coc7	Donor's 30-day cocaine use less than the recipient's maximum number of days could have used in past 30 days
Coc8	Donor's 30-day cocaine use 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/her current age, the donor's cocaine 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first cocaine use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)

Exhibit H.14 (continued)

Constraint #	Constraint
Coc10	If recipient's age at first cocaine use equals his/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/her interview date and date of first drug use (+1) 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/her interview date and birthday (+1)
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/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/her interview date and date of first cocaine use (-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/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 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 crack month user, the past year crack frequency must be less than or equal to the maximum period (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) 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 freq (2) if the recipient is not a past month crack user but a past year crack user, it must be at least 1

Exhibit H.14 (continued)

Constraint #	Constraint
Coc20	(Recipient's proportion of past year crack use * max number of days could have used crack in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Coc21	(Donor's proportion of past year crack use * recipient's number of days could have used crack in past year) greater than or equal to 30-day use
Coc22	Donor's 30-day crack use less than number of days between recipient's interview date and birthday (+1)
Coc23	Donor's 30-day crack use less than the recipient's maximum number of days could have used in past 30 days
Coc24	Donor's 30-day crack use 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/her current age, the donor's crack 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first crack use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1**)**
Coc26	If recipient's age at first crack use equals his/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/her interview date and date of first drug use (+1) 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/her interview date and birthday (+1)
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 (10)
Coc32	If recipient's age at first crack use equals his/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/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/her interview date and date of first crack use (-29)

Exhibit H.15 Cocaine User Restrictions

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
1	(Past month)		Missing		Missing		1	(Coc1-Coc12)
2	(Past month)				Missing		17	(Coc1), (Coc6-Coc9), (Coc11-Coc12)
3	(Past month)		Missing				4	(Coc2-Coc4), (Coc10), (Coc12)
4	(Past year not past month)		Missing				10	(Coc2-Coc4), (Coc12-Coc14)
5	Past year				Missing		24	(Coc6-Coc9), (Coc11-Coc12), (Coc15)
6	Past year		Missing		Missing		1	(Coc2-Coc12), (Coc15)
7	Missing (lifetime use known)		Missing		Missing		71	(Coc2-Coc12), (Coc16-Coc17)
7	Missing (lifetime use imputed)		Missing		Missing		8	
8	(Past month)	(Past month)		Missing		Missing	0	(Coc1), (Coc18-Coc27)
9	(Past month)	(Past month)					0	(Coc1), (Coc22-Coc25), (Coc27-Coc28)

Exhibit H.15 (continued)

Missingness Pattern							Num-ber of Cases	Con-straints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
10	(Past year not missing)	Past year (not missing)		Missing			5	(Coc1), (Coc18-Coc20), (Coc26), (Coc28)
12	(Past month)	Past year				Missing	5	(Coc1), (Coc22-Coc25), (Coc27), (Coc29)
13	(Past month)	Past year		Missing		Missing	0	(Coc1), (Coc18-Coc27), (Coc29)
14	(Past month)	Missing (Lifetime use known)		Missing		Missing	3	(Coc16), (Coc18-Coc26), (Coc30-Coc32)
14	(Past month)	Missing (Lifetime use imputed)		Missing		Missing	0	
15	(Past month)	(Past month)	Missing	Missing			1	(Coc1-Coc4)), (Coc10), (Coc18-Coc20), (Coc26), (Coc28)
16	(Past month)	(Past year but not past month)	Missing	Missing			0	(Coc1-Coc4), (Coc10), (Coc18-Coc20), (Coc26), (Coc33)

Exhibit H.15 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			2	(Coc2-Coc4), (Coc14), (Coc18-Coc20), (Coc33-Coc34)
18	(Past month)	(Past month)			Missing	Missing	1	(Coc1), (Coc6-Coc9), (Coc11), (Coc22-Coc25), (Coc27-Coc28)
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	0	(Coc1-Coc11), (Coc18-Coc28)
20	(Past month)	(Past month)	Missing		Missing	Missing	0	(Coc1-Coc11), (Coc16), (Coc22-Coc25), (Coc27-Coc28)
21	(Past month)	(Past month)		Missing	Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc28)
22	(Past month)	(Past month)	Missing	Missing	Missing		0	(Coc1-Coc11), (Coc18-Coc21), (Coc26), (Coc28)
23	(Past month)	(Past month not past year)	Missing	Missing	Missing		0	(Coc1-Coc11), (Coc18-Coc20), (Coc33), (Coc34)

Exhibit H.15 (continued)

Missingness Pattern							Num-ber of Cases	Con-straints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
24	(Past month)	(Past month)	Missing	Missing		Missing	0	(Coc1-Coc4), (Coc10), (Coc18-Coc26), (Coc28)
25	(Past month)	(Past month)		Missing	Missing		1	(Coc1), (Coc6-Coc9), (Coc18-Coc20), (Coc26), (Coc28)
26	(Past month)	(Past year not past month)		Missing	Missing		0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc 20), (Coc26), (Coc33)
27	(Past month)	(Past month)	Missing			Missing	0	(Coc1-Coc4), (Coc10), (Coc22-Coc25), (Coc27-Coc28)
28	Past year	Past year			Missing	Missing	4	(Coc6-Coc9), (Coc11), (Coc15), (Coc22-Coc25), (Coc27), (Coc29)
29	Past year	Past year	Missing		Missing	Missing	0	(Coc3-Coc11), (Coc15), (Coc21-Coc25), (Coc27), (Coc29)

Exhibit H.15 (continued)

<table>
<tr><th colspan="7">Missingness Pattern</th><th rowspan="2">Num-ber of Cases</th><th rowspan="2">Con-straints</th></tr>
<tr><th>#</th><th>Cocaine Recency</th><th>Crack Recency</th><th>Cocaine 12-Mo. Freq.</th><th>Crack 12-Mo. Freq.</th><th>Cocaine 30-Day Freq.</th><th>Crack 30-Day Freq.</th></tr>
<tr><td>30</td><td>Past year</td><td>Past year</td><td></td><td>Missing</td><td>Missing</td><td>Missing</td><td>1</td><td>(Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc29)</td></tr>
<tr><td>31</td><td>Past year</td><td>Past year</td><td>Missing</td><td>Missing</td><td>Missing</td><td>Missing</td><td>2</td><td>(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc29)</td></tr>
<tr><td>32</td><td>Past year</td><td>Missing (lifetime use known)</td><td></td><td>Missing</td><td>Missing</td><td>Missing</td><td>9</td><td>(Coc1), (Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc30)</td></tr>
<tr><td>32</td><td>Past year</td><td>Missing (lifetime use imputed)</td><td></td><td>Missing</td><td>Missing</td><td>Missing</td><td>0</td><td></td></tr>
<tr><td>33</td><td>Past year</td><td>Missing (lifetime use known)</td><td>Missing</td><td>Missing</td><td>Missing</td><td>Missing</td><td>0</td><td rowspan="2">(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc30), (Coc32)</td></tr>
<tr><td>33</td><td>Past year</td><td>Missing (lifetime use imputed)</td><td>Missing</td><td>Missing</td><td>Missing</td><td>Missing</td><td>0</td></tr>
<tr><td>34</td><td>(Past month)</td><td>Missing (lifetime use known)</td><td></td><td>Missing</td><td>Missing</td><td>Missing</td><td>0</td><td rowspan="2">(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc27), (Coc30), (Coc32)</td></tr>
<tr><td>34</td><td>(Past month)</td><td>Missing (lifetime use imputed)</td><td></td><td>Missing</td><td>Missing</td><td>Missing</td><td>1</td></tr>
</table>

Exhibit H.15 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
35	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	1	(Coc1-Coc11), (Coc18-Coc27), (Coc30)
35	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0	
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	14	(Coc2-Coc11), (Coc16-Coc27), (Coc30)
36	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	5	
36	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	1	
	Lifetime user, nothing missing						5,912	
	Imputed to lifetime nonuse						51	
	Lifetime nonuser, nothing missing						65,609	

Exhibit H.16 Constraints for Hallucinogens (Including LSD and PCP)

Con-straint #	Constraint
Hal1	Donor must be a LSD user (LSD recency not equal to 91)
Hal2	Donor's hallucinogen recency must equal recipient's hallucinogen recency or donor's hallucinogen recency must equal recipient's hallucinogen recency (10)
Hal3	Donor's PCP recency must equal recipient's PCP recency
Hal4	Donor must be a PCP user (PCP recency not equal to 91)
Hal5	Donor's LSD recency must equal recipient's LSD recency
Hal6	Donor must be a LSD and PCP user (LSD and PCP recencies not equal to 91)
Hal7	Donor's must be a past month hallucinogens user (hallucinogen recency = 1)
Hal8	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) 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 (30)
Hal9	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) 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
Hal10	(Recipient's proportion of past year hallucinogen use * max number of days could have used hallucinogens in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Hal11	Donor's 30-day hallucinogen use less than number of days between recipient's interview date and birthday (+1)
Ha12	Donor's 30-day hallucinogen use less than the recipient's maximum number of days could have used hallucinogens in past 30 days

Exhibit H.16 (continued)

Con-straint #	Constraint
Hal13	Donor's 30-day hallucinogen use greater than the recipient's minimum number of days could have used hallucinogens in past 30 days
Hal14	Donor must be a hallucinogen past year (but not past month) or past month user (hallucinogen recency = 1 or 2)
Hal15	Donor must be a LSD past year (but not past month) or past month user (LSD recency = 1 or 2)
Hal16	Donor must be a PCP past year (but not past month) or past month user (PCP recency = 1 or 2)
Hal17	Donor must be a LSD user (LSD recency = 1, 2, or 3)
Hal18	Donor must be a PCP user (PCP recency = 1, 2, or 3)
Hal19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
Hal20	Donor must not be a LSD past year (but not past month) or past month user (LSD recency not equal to 1 or 2)
Hal21	Donor must not be a PCP past year (but not past month) or past month user (PCP recency not equal to 1 or 2)

Exhibit H.17 Hallucinogen User Restrictions (Including LSD and PCP)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
1		Missing (lifetime use known)				6	(Hal1-Hal3)
1		Missing (lifetime use imputed)				0	
2			Missing (lifetime use known)			8	(Hal3), (Hal4-Hal5)
2			Missing (lifetime use imputed)			2	
3		Missing (lifetime use known)	Missing (lifetime use known)			0	(Hal3), (Hal6)
3		Missing (lifetime use imputed)	Missing (lifetime use imputed)			0	
3		Missing (lifetime use known)	Missing (lifetime use imputed)			0	
3		Missing (lifetime use imputed)	Missing (lifetime use known)			0	
4	(Past month)			Missing	Missing	11	(Hal7-Hal13)
5	(Past month)				Missing	19	(Hal7), (Hal11-Hal13)
6	(Past year)			Missing		56	(Hal2-Hal3), (Hal5), (Hal8-Hal10), (Hal14)

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
7	(Past month)	Missing (lifetime use known)			Missing	0	(Hal1), (Hal3), (Hal7), (Hal11-Hal13)
7	(Past month)	Missing (lifetime use imputed)			Missing	0	
8	(Past month)		Missing (lifetime use known)		Missing	0	(Hal4-Hal5), (Hal7), (Hal11-Hal13)
8	(Past month)		Missing (lifetime use imputed)		Missing	0	
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	(Hal6-Hal7), (Hal11-Hal13)
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0	(Hal6-Hal7), (Hal11-Hal13)
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0	
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0	
10	(Past month or Past month not past year)	Missing (lifetime use known)		Missing		0	(Hal1-Hal3), (Hal8-Hal10), (Hal14)
10	(Past month or Past month not past year)	Missing (lifetime use imputed)		Missing		0	

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
11	(Past month or Past month not past year)		Missing (lifetime use known)	Missing		1	(Hal2), (Hal4-Hal5), (Hal8-Hal10), (Hal14)
11	(Past month or Past month not past year)		Missing (lifetime use imputed)	Missing		0	
12	(Past month or Past month not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	(Hal2), (Hal6), (Hal8-Hal10), (Hal14)
12	Past year (not missing)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0	
12	Past year (not missing)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0	
12	(Past month or Past month not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0	
13	(Past month)	Missing (lifetime use known)		Missing	Missing	1	(Hal1), (Hal3), (Hal7-13)
13	(Past month)	Missing (lifetime use imputed)		Missing	Missing	0	

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
14	(Past month)		Missing (lifetime use known)	Missing	Missing	0	(Hal4-Hal5), (Hal7-Hal13)
14	(Past month)		Missing (lifetime use imputed)	Missing	Missing	0	
15	(Past month	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal6-Hal13)
15	(Past month)	Missing (lifetime use imputed)	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 known)	Missing (lifetime use imputed)	Missing	Missing	0	
16	Past year	(Not past month)	(Not past month)		Missing	30	(Hal3), (Hal5), (Hal11-Hal14)
17	Past year	(Not past month)	(Not past month)	Missing	Missing	6	(Hal3), (Hal5), (Hal8-Hal14)
18	Past year	Past year	(Not past month)		Missing	0	(Hal3), (Hal11-Hal15)
19	Past year	(Not past month)	Past year		Missing	0	(Hal5), (Hal11-Hal14), (Hal16)
20	Past year	Past year	Past year		Missing	1	(Hal11-Hall16)
21	Past year	Missing (lifetime use known)	(Not past month)		Missing	32	(Hal3), (Hal11-Hal14), (Hal17)
21	Past year	Missing (lifetime use imputed)	(Not past month)		Missing	0	

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
22	Past year	(Not past month)	Missing (lifetime use known)		Missing	10	(Hal5), (Hal11-Hal14), (Hall18)
22	Past year	(Not past month)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)		Missing	2	(Hal8-Hal14), (Hal17-Hal18)
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0	
24	Past year	Past year	(Not past month)	Missing	Missing	1	(Hal3), (Hal8-Hal15)
25	Past year	(Not past month)	Past year	Missing	Missing	0	(Hal5), (Hal8-Hal14), (Hal16)

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
26	Past year	Past year	Past year	Missing	Missing	0	(Hal8-Hal16)
27	Past year	Missing (lifetime use known)	(Not past month)	Missing	Missing	2	(Hal3), (Hal8-Hal14), (Hal17)
27	Past year	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0	
28	Past year	(Not past month)	Missing (lifetime use known)	Missing	Missing	1	(Hal5), (Hal8-Hal14), (Hal18)
28	Past year	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal8-Hal14), (Hal17-Hal18)
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	
30	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	97	(Hal8-Hal13), (Hal19-Hal21)
30	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	13	

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallucinogen Recency	LSD Recency	PCP Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.		
31	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	101	(Hal8-Hal13), (Hal17), (Hal19), (Hal21)
31	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	4	
31	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0	
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	29	(Hal8-Hal13), (Hal18-Hal20)
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0	
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0	
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	10	(Hal8-Hal13), (Hal17-Hal19)
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	

Exhibit H.17 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	
	Lifetime user, nothing missing					8,135	
	Imputed to lifetime nonuse					226	
	Lifetime nonuser, nothing missing					62,963	

Exhibit H.18 Constraints for Stimulants and Methamphetamines

Constraint #	Constraint
Stm1	Donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year must be less than (or equal) the recipient's maximum possible past year stimulants frequency of use. The recipient's maximum possible stimulants 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 stimulants, as determined by the month of first use (2) if the maximum period the recipient could have used stimulants is greater than 30, but the recipient is a past month stimulants user with a nonmissing 30-day frequency, the past year stimulants 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 stimulants month user, the past year stimulants frequency must be less than or equal to the maximum period (30)
Stm2	Donor's proportion of past year stimulants use * recipient's min number of days could have used stimulants in past year must be greater than (or equal) the recipient's minimum possible past year stimulants frequency of use. The recipient's minimum possible stimulants frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month stimulants user, it must be at least as much as the 30-day freq (2) if the recipient is not a past month stimulants user but a past year stimulants user, it must be at least 1.
Stm3	(Recipient's proportion of past year stimulants use * max number of days could have used stimulants in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Stm4	Donor must be a past month stimulant user (stimulant recency = 1)
Stm5	Donor's meth recency equals the recipient's meth recency
Stm6	If recipient's age at first stimulants use equals his/her current age, (1) recipient's donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Stm7	Donor must be a past year (but not past month) stimulant user (stimulant recency = 2)
Stm8	If recipient's age at first stimulants use equals his/her current age, (1) recipient's donor's proportion of past year stimulants use* recipient's max number of days could have used stimulants in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (-29) and (2) donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than the recipient's days between his/her interview date and birthday (-29)
Stm9	Donor must be a past month or past year (but not past month) stimulant user (stimulants recency = 1 or 2)

Exhibit H.18 (continued)

Constraint #	Constraint
Stm10	If recipient's age at first stimulants use equals his/her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first stimulants use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Stm11	Donor's stimulants recency must equal recipient's stimulants recency or donor's stimulants recency must equal recipient's stimulants recency (10).
Stm12	Donor must be a past month, past year (but not past month), or lifetime (but not past year) meth user (meth recency = 1, 2, or 3)
Stm13	If the number of days between the recipient's interview and birthday (+1) is between 0 and 30, meth recency must not equal 2 or 3
Stm14	If the number of days between the recipient's interview and birthday (+1) is between 0 and 365, meth recency must not equal 3
Stm15	If recipient's age at first stimulants use equals his/her current age or the recipient's age at first meth use equals his/her current age or the recipient's number of days between his/her interview date and date at first meth use less than 30, the donor's recency must not equal 3
Stm16	If recipient's age at first stimulants use equals his/her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first stimulants use (-29) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (-29)
Stm17	Donor must be a past month or past year (but not past month) meth user (meth recency = 1 or 2)

Exhibit H.19 Stimulants User Restrictions (Includes Methamphetamines)

Missingness Pattern					
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency	Number of Cases	Constraints
1	(Past month)		Missing	52	(Stm1-Stm6)
2	(Past year but not past month)		Missing	154	(Stm1-Stm3), (Stm5), (Stm7-Stm8)
3	Past year			1	(Stm5), (Stm8-Stm10)
4	Past year		Missing	12	(Stm1-Stm3), (Stm5-Stm6), (Stm8-Stm9)
5	Missing (lifetime use known)		Missing	107	(Stm1-Stm3), (Stm5-Stm6), (Stm8)
5	Missing (lifetime use imputed)		Missing	9	
6		Missing (lifetime use known)		2	(Stm11-Stm15)
6		Missing (lifetime use imputed)		3	
7	(Past month)	Missing (lifetime use known)	Missing	1	(Stm1-Stm4), (Stm6), (Stm12-Stm15)
7	(Past month)	Missing (lifetime use imputed)	Missing	0	
8	(Past year not past month)	Missing (lifetime use known)	Missing	0	(Stm1-Stm3), (Stm7-Stm8), (Stm12-Stm15)
8	(Past year not past month)	Missing (lifetime use imputed)	Missing	0	
9	Past year	Missing (lifetime use known)		7	(Stm9-Stm10), (Stm12-16)
9	Past year	Missing (lifetime use imputed)		0	
10	Past year	Missing (lifetime use known)	Missing	1	(Stm1-Stm3), (Stm6), (Stm8-Stm9), (Stm12-Stm15)
10	Past year	Missing (lifetime use imputed)	Missing	0	
11	Past year (not missing)	Past year		0	(Stm11), (Stm13), (Stm17)

Exhibit H.19 (continued)

Missingness Pattern				Number of Cases	Constraints
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency		
	Past month	Past year	Missing	0	(Stm1-Stm4), (Stm6), (Stm9), (Stm13)
13	(Past year not past month)	Past year	Missing	0	(Stm1-Stm3), (Stm7-Stm8), (Stm13), (Stm17)
14	Past year	Past year		1	(Stm9-Stm10), (Stm13), (Stm16-Stm17)
15	Past year	Past year	Missing	0	(Stm1-Stm3), (Stm6), (Stm8-Stm9), (Stm17)
16	Missing (lifetime use known)	Missing (lifetime use known)	Missing	45	(Stm1-Stm3), (Stm6), (Stm8), (Stm12-Stm15)
16	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	4	
16	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	0	
	Lifetime user, nothing missing			3,962	
	Imputed to lifetime nonuse			159	
	Lifetime nonuser, nothing missing			67,244	

H.2.3 Health Insurance

IRINSUR (overall health insurance 1999 method) and IRINSUR2 (overall health insurance 2000 method) were created separately, each with an accompanying version of IRPINSUR (private health insurance). Hence, two multivariate imputations were performed: IRINSUR-IRPINSUR and IRINSUR2-IRPINSUR. The version of IRPINSUR that was created with IRINSUR2 was the private health insurance variable released for analysis. However, since the imputations were separate for the 1999 and 2000 methods of determining overall health insurance, separate tables are given below. In order to maintain consistency between the three variables IRINSUR, IRINSUR2, and IRPINSUR, the imputed value of IRINSUR was changed from "no" to "yes" in 70 cases, and was changed from "yes" to "no" in 83 cases.

Exhibit H.20A Health Insurance (2000 Method)

Missingness Pattern			Number of Cases	Logical Constraints
#	Overall Health Insurance	Private Health Insurance		
1[1]	Missing		103	None
2	Missing	Missing	761	None
3		Missing	145	None

[1] This pattern only occurs if the response to the private health insurance question is "no". Obviously, if the response to the private health insurance question is “yes”, the overall health insurance response would logically also be "yes."

Exhibit H.20B Health Insurance (1999 Method)

Missingness Pattern			Number of Cases	Logical Constraints
#	Overall Health Insurance	Private Health Insurance		
1[1]	Missing		94	None
2	Missing	Missing	799	None
3		Missing	107	None

[1] This pattern only occurs if the response to the private health insurance question is "no". Obviously, if the response to the private health insurance question is “yes”, the overall health insurance response would logically also be "yes."

H.2.4 Source of Income

There are a large number of missingness patterns for the source of income variables because they are imputed together in a set. The only logical constraint applied to the potential donors is 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.

H.3 Exhibits Showing Missingness Patterns and the Portions of the Predictive Mean Vector Used in the Calculation of the Mahalanobis Distance (with Adjustments)

H.3.1 Drug Lifetime Use

There are a large number of missingness patterns for drug lifetime use. The response to the gate question for cigarettes must be nonmissing for the survey to be considered complete, but any combination of the other lifetime drug variables may be missing. There are 14 other gate questions in the questionnaire, plus several subgate questions.

The probabilities associated with the 14 gate questions (**Exhibit H.21**) form the full predictive mean vector. Only the probabilities associated with the gate questions for which the responses are missing are used in the predictive mean vector for each item nonrespondent.

Exhibit H.21 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 Lifetime	P(Lifetime User)
Hallucinogens/LSD/PCP 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)

Section H.3.2 Drug Recency and Frequency

Exhibit H.22 Portion of the Predictive Mean Vector for Cigarette Users

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	30-Day Frequency		
1	Past year	Missing	15	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)
2	Missing	Missing	57	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM
3	(Past month)	Missing	61	1. D 2. PM
4	Not past year		363	1. R3/(R3+R4)
5	Not past month		324	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

[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. D = P(daily use | past month use)
5. PM = P(use on a given day in the past month | past month use)

Exhibit H.23 Portion of the Predictive Mean Vector for Cigar Users

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	30-Day Frequency		
1	Past year	Missing	17	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
2	Missing	Missing	42	1. R1 2. R2 3. R3 4. R1*PM
3	(Past month)	Missing	20	1. PM
4	Not past year		258	1. R3/(R3+R4)
5	Not past month		364	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

[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)

Exhibit H.24 Portion of the Predictive Mean Vector for Smokeless Tobacco Users

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Predictive Mean Vector[1]
1	(Past month)	(Past month)	Missing	Missing	2	1. DC 2. PMC 3. DS 4. PMS
2	(Past month)		Missing		4	1. DC 2. PMC
3		(Past month)		Missing	4	1. DS 2. PMS
4		Lifetime		Missing	5	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
5	(Past month)	Lifetime	Missing	Missing	0	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
6	Lifetime		Missing		13	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
7	Lifetime	(Past month)	Missing	Missing	0	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS
8		Past year		Missing	2	No cases
9	Past year		Missing		9	1. R1/(R1+R2) 2. RC1*DC/(RC1+RC2) 3. RC1*(1-DC)*PMC/(RC1+RC2)

Exhibit H.24 (continued)

Missingness Pattern					Num-ber of Cases	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.		
10	Lifetime	Lifetime	Missing	Missing	3	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
11	Not past year				84	1. R3/(R3+R4)
12		Not past year			45	1. R3/(R3+R4)
13	Not past year	Not past year			9	1. R3/(R3+R4)
14	Not past month				138	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
15		Not past month			46	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
16	Not past month	Not past month			15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
17	Not past month	(Past month)		Missing	0	No cases
18	(Past month)	Not past month	Missing		0	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC
19	Not past month	Lifetime		Missing	0	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS

Exhibit H.24 (continued)

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Predictive Mean Vector[1]
20	Lifetime	Not past month	Missing		0	No cases
21	Not past month	Not past year			0	No cases
22	Not past year	Not past month			1	

[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)

Exhibit H.25 Portion of the Predictive Mean Vector for Alcohol Users

Missingness Pattern					Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Freq.	30-Day Freq.	30-Day Binge Drink		
1	(Past month)	Missing	Missing		20	1. PM 2. PY
2	(Past month)		Missing		275	1. PM
3	(Past month)	Missing			170	1. PY
4	(Past year but not past month)	Missing			220	1. PY
5	(Past month)			Missing	938	1. PMB
6	(Past month)		Missing	Missing	11	1. PM 2. PMB
7	(Past month)	Missing		Missing	104	1. PY 2. PMB
8	(Past month)	Missing	Missing	Missing	46	1. PM 2. PY 3. PMB
9	Past year		Missing	Missing	485	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)
10	Past year	Missing	Missing	Missing	62	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)
11	Lifetime	Missing	Missing	Missing	515	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB

[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)

Exhibit H.26 Portion of the Predictive Mean Vector for Marijuana Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	20	1. PM 2. PY
2	(Past month)		Missing	8	1. PM
3	(Past month)	Missing		61	1. PY
4	(Past year but not past month)	Missing		68	1. PY
5	Past year		Missing	94	1. R1/(R1+R2) 2. R1*PM/(R1*R2)
6	Past year	Missing	Missing	97	1. R1/(R1+R2) 2. R1*PM/(R1*R2) 3. PY
7	Lifetime	Missing	Missing	324	1. R1 2. R2 3. R1*PM 3. (R1+R2)*PY

[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)

Exhibit H.27 Portion of the Predictive Mean Vector for Cocaine Users

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
1	(Past month)		Missing		Missing		1	1. PM 2. PY
2	(Past month)				Missing		17	1. PM
3	(Past month)		Missing				4	1. PY
4	(Past year not past month)		Missing				10	1. PY
5	Past year				Missing		24	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year		Missing		Missing		1	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing		Missing		Missing		79	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
8	(Past month)	(Past month)		Missing		Missing	0	1. PM 2. PY
9	(Past month)	(Past month)				Missing	0	1. PM
10	(Past month)	(Past month)		Missing			5	no cases
11[2]	(Past year not missing)	(Past year but not past month)		Missing			0	1. PY
12	(Past month)	Past year				Missing	5	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
13	(Past month)	Past year		Missing		Missing	0	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
14	(Past month)	Lifetime		Missing		Missing	3	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Exhibit H.27 (continued)

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
15	(Past month)	(Past month)	Missing	Missing			1	No cases
16	(Past month)	(Past year but not past month)	Missing	Missing			0	1. PY
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			2	1. PY
18	(Past month)	(Past month)			Missing	Missing	1	1. PM
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	0	1. PM 2. PY
20	(Past month)	(Past month)	Missing		Missing	Missing	0	No cases
21	(Past month)	(Past month)		Missing	Missing	Missing	0	No cases
22	(Past month)	(Past month)	Missing	Missing	Missing		0	1. PM 2. PY
23	(Past month)	(Past year not past year)	Missing	Missing	Missing		0	1. PM 2. PY
24	(Past month)	(Past month)	Missing	Missing		Missing	0	No cases
25	(Past month)	(Past month)		Missing	Missing		1	No cases
26	(Past month)	(Past year not past month)		Missing	Missing		0	No cases
27	(Past month)	(Past month)	Missing			Missing	0	No cases
28	Past year	Past year			Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2)

Exhibit H.27 (continued)

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
29	Past year	Past year	Missing		Missing	Missing	0	No cases
30	Past year	Past year		Missing	Missing	Missing	1	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
31	Past year	Past year	Missing	Missing	Missing	Missing	2	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
32	Past year	Lifetime		Missing	Missing	Missing	9	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
33	Past year	Lifetime	Missing	Missing	Missing	Missing	0	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
34	(Past month)	Lifetime		Missing	Missing	Missing	1	No cases
35	(Past month)	Lifetime	Missing	Missing	Missing	Missing	1	No cases
36	Lifetime	Lifetime	Missing	Missing	Missing	Missing	20	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Note: Includes crack users, and cocaine users who were not crack users

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month cocaine use | lifetime cocaine use)
2. R2 = P(past year but not past month cocaine use | lifetime cocaine use)
3. PM = P(cocaine use on a given day in the past month | past month use of cocaine)
4. PY = P(cocaine use on a given day in the past year | past year use of cocaine)

[2]Due to a programming error, the three respondents fitting missingness pattern #11 were misclassified under missingness pattern #10. As a result, the donors assigned to these three respondents were all past month users, and the imputed 12-month frequency for crack might have been slightly affected.

Exhibit H.28 Portion of the Predictive Mean Vector for Heroin Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	0	No cases
2	(Past month)		Missing	0	1. PM
3	(Past month)	Missing		0	No cases
4	(Past year not past month)	Missing		0	1. PY
5	Past year		Missing	7	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Lifetime	Missing	Missing	18	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

[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)

Exhibit H.29 Portion of the Predictive Mean Vector for Hallucinogen Users

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
1		Lifetime				6	1. R1 2. R2
2			Lifetime			10	1. R1 2. R2
3		Lifetime	Lifetime			0	No cases
4	(Past month)			Missing	Missing	11	1. PM 2. PY
5	(Past month)				Missing	19	1. PM
6	(Past year)			Missing		56	1. PY
7	(Past month)	Lifetime			Missing	0	No cases
8	(Past month)		Lifetime		Missing	0	No cases
9	(Past month)	Lifetime	Lifetime		Missing	0	1. R1 2. R2 3. PM
10	(Past year)	Lifetime		Missing		0	No cases
11	(Past year)		Lifetime	Missing		1	No cases
12	(Past year)	Lifetime	Lifetime	Missing		0	No cases
13	(Past month)	Lifetime		Missing	Missing	1	No cases
14	(Past month)		Lifetime	Missing	Missing	0	1. R1 2. R2 3. PM 4. PY
15	(Past month)	Lifetime	Lifetime	Missing	Missing	0	No cases
16	Past year	(Not past month)	(Not past month)		Missing	30	1. R1/(R1+R2) 2. R1*PM/(R1+R2)

Exhibit H.29 (continued)

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
17	Past year	(Not past month)	(Not past month)	Missing	Missing	6	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
18	Past year	Past year	(Not past month)		Missing	0	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
19	Past year	(Not past month)	Past year		Missing	0	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
20	Past year	Past year	Past year		Missing	1	No cases
21	Past year	Lifetime	(Not past month)		Missing	32	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
22	Past year	(Not past month)	Lifetime		Missing	10	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
23	Past year	Lifetime	Lifetime		Missing	2	No cases
24	Past year	Past year	(Not past month)	Missing	Missing	1	No cases
25	Past year	(Not past month)	Past year	Missing	Missing	0	No cases
26	Past year	Past year	Past year	Missing	Missing	0	No cases
27	Past year	Lifetime	(Not past month)	Missing	Missing	2	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
28	Past year	(Not past month)	Lifetime	Missing	Missing	1	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
29	Past year	Lifetime	Lifetime	Missing	Missing	0	No cases
30	Lifetime	(Not past year)	(Not past year)	Missing	Missing	107	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Exhibit H.29 (continued)

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
31	Lifetime	Lifetime	(Not past year)	Missing	Missing	105	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
32	Lifetime	(Not past year)	Lifetime	Missing	Missing	29	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
33	Lifetime	Lifetime	Lifetime	Missing	Missing	10	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Note: Hallucinogen users include users of LSD and PCP.

[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)

Exhibit H.30 Portion of the Predictive Mean Vector for Inhalant Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	4	1. PM 2. PY
2	(Past month)		Missing	4	1. PM
3	(Past month)	Missing		6	1. PY
4	(Past year not past month)	Missing		23	1. PY
5	Past year		Missing	28	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Lifetime	Missing	Missing	257	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

[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)

Exhibit H.31 Portion of the Predictive Mean Vector for Users of Pain Relievers, Tranquilizers, and Sedatives

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency		
1	(Past month)	Missing	PR:47 TR:7 SD:1	1. PY
2	(Past year not past month)	Missing	PR:47 TR:18 SD:4	1. PY
3	Past year		PR:3 TR:1 SD:1	1. R1/(R1+R2)
4	Past year	Missing	PR:15 TR:5 SD:1	1. R1/(R1+R2) 2. PY
5	Lifetime	Missing	PR:345 TR:96 SD:36	1. R1 2. R2 3. (R1+R2)*PY

Note: The missingness patterns and predictive mean vectors for the pain relievers, tranquilizers, and sedatives modules were identical. When required, the identifiers "PR," "TR," and "SD" are used for pain relievers, tranquilizers, and sedatives, respectively.

[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)

Exhibit H.32 Portion of the Predictive Mean Vector for Stimulant Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency		
1	(Past month)		Missing	52	1. PY
2	(Past year not past month)		Missing	154	1. PY
3	Past year			1	1. R1/(R1+R2)
4	Past year		Missing	12	1. R1/(R1+R2) 2. PY
5	Lifetime		Missing	116	1. R1 2. R2 3. (R1+R2)*PY
6		Lifetime		5	1. R1 2. R2
7	(Past month)	Lifetime	Missing	1	1. R1 2. R2 3. PY
8	(Past year not past month)	Lifetime	Missing	0	No cases
9	Past year	Lifetime		7	1. R1/(R1+R2)
10	Past year	Lifetime	Missing	1	No cases
11	(Past year)	Past year		0	No cases
12	(Past month)	Past year	Missing	0	No cases
13	(Past year not past month)	Past year	Missing	0	No cases
14	Past year	Past year		1	1. R1/(R1+R2)
15	Past year	Past year	Missing	0	1. R1/(R1+R2) 2. PY
16	Lifetime	Lifetime	Missing	49	1. R1 2. R2 3. (R1+R2)*PY

Note: Users of stimulants include 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.3.3 Health Insurance

IRINSUR (overall health insurance 1999 method) and IRINSUR2 (overall health insurance 2000 method) were created separately, each with an accompanying version of IRPINSUR (private health insurance). Hence, two multivariate imputations were performed: IRINSUR-IRPINSUR and IRINSUR2-IRPINSUR. Separate predictive mean vectors were used for the two multivariate imputations. Although the predicted mean from the private health insurance model was the same for both imputations, two models were fit for overall health insurance, one for the 1999 method and one for the 2000 method. The predictive mean corresponding to overall health insurance varied according to whether IRINSUR or IRINSUR2 was being imputed.

The version of IRPINSUR that was created with IRINSUR2 was the private health insurance variable released for analysis. However, since the imputations were separate for the 1999 and 2000 methods of determining overall health insurance, separate tables are given below. In order to maintain consistency between the three variables IRINSUR, IRINSUR2, and IRPINSUR, the imputed value of IRINSUR was changed from "no" to "yes" in 70 cases, and was changed from "yes" to "no" in 83 cases.

Exhibit H.33A Health Insurance (2000 Method)

Missingness Pattern					
#	Overall Health Insurance	Private Health Insurance	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Missing	(No)	103	None	1. OVR
2	Missing	Missing	761	None	1. OVR 2. PRV
3	(Yes)	Missing	145	None	1. PRV

[1]The predictive mean vector elements are defined by the following:
1. OVR = P(overall insurance) 2. PRV = P(private insurance)

Exhibit H.33B Health Insurance (1999 Method)

Missingness Pattern					
#	Overall Health Insurance	Private Health Insurance	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Missing	(No)	94	None	1. OVR
2	Missing	Missing	799	None	1. OVR 2. PRV
3	(Yes)	Missing	107	None	1. PRV

[1]The predictive mean vector elements are defined by the following:
1. OVR = P(overall insurance) 2. PRV = P(private insurance)

H.3.4 Source of Income

There are a large number of missingness patterns for the source of income variables. The probabilities associated with each of the family source of income variables form the full predictive mean vector. (For months-on-welfare, the probability used is the probability of receiving welfare payments or services in a given month in the past 12 months.) Only the probabilities associated with the family-level variables for which the responses are missing are used in the predictive mean vector for each item nonrespondent.

Appendix I: Quality Control Procedures Used in Drug Use Imputations

Appendix I: Quality Control Procedures Used in Drug Use Imputations

The imputation process for drug use imputations occurs in three basic steps, and quality control checks are done at each of the three steps. In addition to the checks listed below, all SAS programs run by members of the imputation team are reviewed by at least two team members for obvious errors, such as messages in the SAS log, model convergence, and missing values.

Step 1. Adjust Weights for Item Nonresponse to Be Used in Models

In this step, it is necessary to define a set of variables over which item nonresponse is defined. To be classified as a "complete" respondent, a respondent would have to respond to all the questions within the variable set. Only complete respondents would be used to build the models in the next step. As a general practice, the weights are adjusted so that the weights for complete respondents represent the entire domain, where "domain" is defined as the population of interest (e.g., lifetime users who are 12 to 17 years of age). This is accomplished this using an item response propensity model, where the predicted value that comes out of the model is given by

P (survey respondent is an item respondent|respondent is in the domain of interest).

The inverse of this probability is multiplied by the weights.

- The output of the response propensity modeling program is checked for singularities. Any singularities that occur are investigated, and the model is corrected.

- An indicator is calculated in the response propensity program that measures the maximum adjustment to the weights. In most cases, the adjusted weights should look very much like the original weights. If the maximum adjustment is too high, it is likely that the adjustment is doing more harm than good, probably due to a very bad model. Large maximum adjustments are investigated, so that any final adjustment that is applied is reasonable.

- The number of people identified as item nonrespondents is recorded. This number should be the same as the number of people excluded from the model-building process.

- Using PROC MEANS, the sum totals for the independent variables are compared before and after the adjustment. If these sums are equal, the adjustment procedures worked.

- The output is checked for missing values.

- Any changes to existing programs are checked by other members of the imputation team.

Step 2. Predictive Mean Modeling

For each question, modeling procedures are used to determine the predicted mean values for each respondent. For example, a model is used to determine the probability of lifetime usage of a given drug based on the responses to the gate question. Predicted mean values are determined regardless of whether the respondent answered the question. These predicted means are calculated based on binomial and multinomial logistic models, Poisson regression models, as well as ordinary weighted least squares regression models, with the response variable appropriately transformed.

- Many of the independent variables are categorical variables and are converted into a set of indicator variables in an intermediate step. A list of a few observations on the dataset is printed to ensure that all of the indicator variables were created correctly.

- Convergence is ensured, and the output is checked to verify that everything worked in the regression model.

- A subset of observations on the output dataset is investigated more closely to ensure that all of the predicted values and indicators make sense.

- For age at first use, the predicted age at first use is crossed with the respondent's age. It is possible for the *predicted* age at first use to exceed the respondent's age, but this would indicate severe problems with the model. Such situations are investigated.

- Any changes to existing programs are checked by other members of the imputation team.

Step 3. Final Assignment of Imputed Values

The predicted means from Step 2 are used to determine the final assignments of imputed values in a hot-deck step. The goal of this step is to make donors and recipients as alike as possible. A neighborhood of potential donors is used, if possible, so that the donor selected is different each time the procedure is run. However, all potential donors in a neighborhood must have very similar predicted means. Quality control checks in this step have two objectives: (1) to ensure that the imputed values are consistent with preexisting nonmissing values, and (2) to ensure that the imputed values were assigned as intended.

- At the end of the hot-deck assignment program, two lists are created and entail (a) any missing values (cases where the program was unsuccessful in assigning an imputed value), and (b) any cases where the imputed value is not consistent with preexisting nonmissing values. The cases printed out in (b) are cases where one or more variables were imputed, and one or more of these variables violate one or more of the following conditions:

    - ☐ The 12-month frequency must be equal to or exceed 30-day frequency.

    - ☐ Past month users must have a valid 30-day frequency (not a skip code).

    - ☐ Past year users must have a valid 12-month frequency (not a skip code).

    - ☐ For alcohol, 30-day frequency must exceed or equal "binge" drinking frequency.

    - ☐ For parent-child drugs (e.g., cocaine and crack, smokeless tobacco and snuff), the parent drug recency must be no later than the child drug's recency.

    - ☐ For cocaine and crack, the cocaine 12-month frequency must equal or exceed the crack 12-month frequency, if it exists.

    - ☐ For cocaine and crack, the cocaine 30-day frequency must equal or exceed the crack 30-day frequency, if it exists.

    - ☐ The recency and frequency of use variables that are imputed must be consistent with the time period between the birthday and interview date, as well as the time period between the interview date and the month that the respondent began using, if that variable is available. For example, if the respondent is not a past month user, the imputed 12-month frequency of use cannot exceed the maximum usage period minus 30.

    - ☐ If the respondent's age is equal to the age at first use, the recency of use must be imputed to be past month or past year not past month.

- ☐ For past month users, the 30-day frequency must exceed the 12-month frequency minus 335.

- ☐ If the edited age at first use is equal to the current age of the respondent, the imputed recency must be consistent with the time period between the birthday and the interview date, and it must be consistent with the month that the respondent began using.

- ☐ The age at first use must not exceed the respondent's age.

- ☐ For parent-child drugs, the parent drug's age at first use must be at least as early as the child's drug age at first use.

- ☐ The respondent's age at first use must not equal the respondent's age if the recency is "not in the past year."

- Only looking at cases where some imputation is required, the distribution of the imputed values is compared with the distribution of nonimputed values. Unusual patterns in these distributions are investigated. This includes the distribution of lifetime users versus nonlifetime users, the distributions of recency and frequency of use, and the age at first use distributions.

- Looking at all respondents, the distribution of values is regarded after imputation has been implemented.

- For multivariate imputations, each pattern of missingness is treated separately. The distribution of imputed values within each missingness pattern is investigated. For example, if it is known that a respondent is a past year user, one would expect both past month and past year users, not just past month users, among the imputed values.

- The imputed values are crossed with the imputation indicators to ensure that the indicators were created correctly.

- In the multivariate imputation of recency and frequency of use, provisional imputed values are used in the process before a final vector of predicted means is created. The provisional imputed recencies are crossed with the final imputed recencies. If something went wrong in the final multivariate hot-deck step, this check is set up to identify it.

- The distribution of imputed values in 1999 is compared with the distribution of imputed values in 2000.

- The assigned date of first use should be consistent with the given birth date and the imputation-revised age at first use.

- The assigned date of first use should be consistent with the given interview date and the imputation-revised recency/frequency of use variables.

- Respondents failing either of the two preceding checks are carefully examined. Occasionally, the error is unavoidable. This happens when the age at first use, recency of use, and interview date are inconsistent by only 1 day, even after editing. In particular, this can occur if the birthday or interview date occur on the first of the month. It is important to ensure that all inconsistencies that appear are of this type.

- The imputation-revised recency is crossed with the imputation-revised age at first use to ensure that all "never used" codes are assigned correctly and that all users have a valid age at first use.

- The imputation-revised year and month of first use are crossed with the edited year and month of first use to ensure that all valid edited year/months are being carried over to the imputation-revised year/month of first use.

- A frequency of the imputation-revised month/day/year of first use variables is run to ensure that all are within the acceptable numbers (i.e. month is between 1 and 12 or 99 for never used).

- PROC MEANS is run to check for missing values.

- Sometimes an error is discovered further along in the process, so that a patch is necessary for earlier imputations. When the variables are reimputed and the dataset updated, it is crucial to compare the old (incorrect) imputation-revised variable and the new corrected variable with the reimputed values. This is necessary to ensure that (1) the changes made were within expectation, and that (2) other cases did not inadvertently change with the correction. Cases that have unanticipated changes should be investigated individually.

- For some of the programs, any changes to existing programs are checked by other members of the imputation team.