2002 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

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

Deliverable No. 28

Project Director: Thomas G. Virag

Authors:

Eric A. Grau
Kortnee Barnett-Walker
Elizabeth Copello
Peter Frechtel
Amy Licata
Bing Liu
Peilan Martin
Dawn M. Odom

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

February 2004

2002 NATIONAL SURVEY ON DRUG USE AND HEALTH

IMPUTATION REPORT

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

Deliverable No. 28

Authors:

Eric A. Grau
Kortnee Barnett-Walker
Elizabeth Copello
Peter Frechtel
Amy Licata
Bing Liu
Peilan Martin
Dawn M. Odom

Project Director:
Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, MD 20857

Prepared by:

RTI International
Research Triangle Park, NC 27709

February 2004

Acknowledgments

This report would not be possible without the guidance and input of staff from the Office of Applied Studies (OAS). In particular, useful comments were provided by Dicy Painter, Joe Gfroerer, and Art Hughes. Special thanks are also due to several 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 9.5.3 and 9.5.4. Jill Webster provided valuable input in the writing of the immigrant status variables in Section 5.2. Finally, Lisa Stallings and Scott Chestnut 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)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Table of Contents (continued)

Section **Page**

Appendix

List of Tables

Page

List of Exhibits

1. Introduction

The 2002 National Survey on Drug Use and Health (NSDUH)[1] was implemented using a 50-State multistage cluster design. This design has been in use since the 1999 survey, when this survey was called the National Household Survey on Drug Abuse (NHSDA). Other major changes in the 1999 survey from surveys in previous years included the introduction of computer-assisted interviewing (CAI) methods for both screening households and interviewing selected respondents. An interview using paper-and-pencil interviewing (PAPI) methods also was included in the 1999 survey for consistency with previous years. However, in the surveys after the 1999 one, only a CAI sample was selected. The 50-State design was developed for the 1999 survey to allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide direct estimates for eight large States and estimates based on small area estimation (SAE) methods for the remaining States and the District of Columbia. This resulted in a major increase in sample size at the national level (from about 20,000 to 67,500 per year).

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

Although the design of the 2002 survey was similar to the design of the 1999 through 2001 surveys, there were important methodological differences in the 2002 NSDUH that affected the 2002 NSDUH estimates. In addition to the name change for the 2002 survey, each NSDUH respondent for this survey was given an incentive payment of $30. Also, information from the 2000 decennial census figures was used for the first time in the 2002 NSDUH weighting procedures.

This report focuses on the imputation procedures implemented for the 2002 NSDUH. The eligibility and completeness criteria are discussed in Chapter 2, followed by a summary of the implemented imputation procedures in Chapter 3. Chapters 4 and 5 describe the imputation procedures applied to the core and noncore demographic variables, respectively. The drug imputation procedures are discussed in Chapter 6. The imputation procedures for nicotine dependence differed from those used for other variables, and are described in Chapter 7. Most of the editing procedures that were applied to the demographic, drug, nicotine dependence, and

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

health insurance variables discussed in Chapters 4, 5, 6, and 7 are summarized by Kroutil (2003a, 2003b, 2003c). The editing procedures for the income and household composition variables, however, are discussed in this document. Chapter 8 describes the edits applied to the household roster, the creation and imputation of missing values in the roster-derived household composition variables, and the creation of respondent-level variables with individual roster information. Chapter 9 summarizes the editing and imputation procedures applied to the income variables. A new approach was used with the imputation of health insurance variables for the 2002 NSDUH. In particular, missing values in the constituent variables for overall health insurance were imputed for the first time. Procedures for the imputation of missing values in the health insurance variables are described in Chapter 10.

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 methodology developed specifically for the NSDUH, 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 and Hispanicity is described in Appendix D. (Coding of alpha-specify responses for other variables is summarized by Kroutil, 2003a, 2003b, 2003c.) The covariates in each of the imputation models are listed in Appendix E. A summary of the number of respondents who met various constraints that could be loosened in the imputation process is provided in Appendix F. Appendix G gives details of the vector of predictive 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 measures used in the imputation procedures are summarized in Appendix H. Reasons that interviewers gave for overriding consistency checks in the household roster are presented in Appendix I, along with evaluations of their legitimacy and the resulting actions in the editing of the roster. For the 2002 NSDUH questionnaire specifications for programming, refer to RTI (2003).

2. Eligibility and Completeness Rules

2.1 Eligibility Criteria

The population of eligible respondents for the 2002 National Survey on Drug Use and Health (NSDUH)[2] was all civilian, noninstitutionalized residents of the United States (including the District of Columbia) aged 12 or older. As in other recent NSDUHs, this population included residents of noninstitutional group quarters (e.g., homeless shelters, rooming houses, dormitories, and group homes), and civilians residing on military bases. Persons excluded from the 2002 survey included those with no fixed household address (e.g., homeless transients *not* in shelters), residents of institutional group quarters, (e.g., jails and hospitals), 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. However, due to screening errors, some individuals were selected, but later were determined to be ineligible at the time of interview. For a summary of the number of eligible persons rostered and the completed interviews obtained in the 2002 NSDUH, see Table 2.1.

Table 2.1 Household and Person Eligibility and Response Rates, 2002 NSDUH

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases
CAI*	178,013	150,162	136,349	284,443	80,581	68,225	68,126

* CAI = computer-assisted interviewing.

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 paper-and-pencil interviewing (PAPI) questionnaire in 1999 and NSDUHs prior to 1999, no logical inference could be made from information within a section if the gate question was not answered. This was due to the fact that the computer-assisted interviewing (CAI) instrument routed respondents out of a section if the gate question was not answered. For a summary of the number of completed cases in the 2002 NSDUH, see Table 2.1.

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

3. Overview of Item Imputation Procedures

3.1 Introduction

As with most large-scale sample surveys, the 2002 National Survey on Drug Use and Health (NSDUH)[3] faced the problem of analyzing datasets that contained missing responses for some items. In association with this there were other issues, such as inconsistent or invalid responses and violation of skip patterns. Although the instrument was designed to enforce skip patterns, which has reduced inconsistencies relative to the paper-and-pencil interview (PAPI), and perform some consistency checks, inconsistent and invalid responses still occurred. These response errors were an obvious source of bias that were considered in the analysis of NSDUH data (Cox & Cohen, 1985).

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

Since the 1999 survey, the NSDUH has been conducted using computer-assisted interviewing (CAI) methods, and the CAI instrument has been the only version used since the 2000 survey. To maintain consistency with NSDUHs since 1999, most of the procedures in the 2002 sample were identical to those used in the 1999 (CAI), 2000, and 2001 surveys. Each year, however, minor modifications were made to the instrument, which subsequently required adjustments to the imputation procedures, and the 2002 NSDUH was no exception. As in the 2001 NHSDA, the procedure developed specifically for the 1999 survey, predictive mean neighborhoods (PMN), was applied to most of the variables requiring imputation in the 2002 survey. Exceptions to this rule included imputations for nicotine dependence, which was also handled differently in the 2001 NHSDA, and the immigrant variables. Exhibit 3.1 provides a brief summary of the types of imputation procedures used for each of the variables imputed in the samples from 1999 to 2002. This chapter provides a brief description of PMN, the imputation procedure most used in the 2002 NSDUH, followed by a description of the other procedures used in the 2002 NSDUH, and a summary of the changes in imputation procedures from the 2001 NHSDA to 2002 NSDUH.

---

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

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

Variable	**1999**[1]	**2000**	**2001**	**2002**
Interview Date	Random[2]	Random	None	None
Age	None[3]	None	None	None
Birth Date	None	Random	Random	Random
Gender	None	None	None	None
Race	UHD[4]	MPMN[5]	MPMN[5]	MPMN[5]
Hispanic-Origin Indicator	UHD	UPMN[6]	UPMN[6]	UPMN[6]
Marital Status	UHD	MPMN	MPMN	MPMN
Hispanic-Origin Group	UHD	MPMN	MPMN	MPMN
Education	UHD	UHD	MPMN	MPMN
Employment Status	UHD	UHD	MPMN	MPMN
Immigrant Variables	Not imputed	Not imputed	Not imputed	WHD[7]
Health Insurance	MPMN	MPMN	MPMN	MPMN[8]
Drug Lifetime Usage (enters into recency)	UPMN	MPMN	MPMN	MPMN
Drug Recency of Use	MPMN	MPMN	MPMN	MPMN
Drug Frequency of Use (12 months)	MPMN	MPMN	MPMN	MPMN
Drug Frequency of Use (30 days)	MPMN	MPMN	MPMN	MPMN
Binge Drinking[9] Frequency (30 days)	MPMN	MPMN	MPMN	MPMN
Age at First Use	UPMN	UPMN	UPMN	UPMN
Age at First Daily Cigarette Use	UPMN	UPMN	UPMN	UPMN
Personal and Family Income Binary Variables	MPMN	MPMN	MPMN	MPMN
Personal and Family Income Finer Categories	UPMN	UPMN	UPMN	UPMN
Nicotine Dependence	Not imputed	Not imputed	Regression	Regression
Household Size (Roster-Derived Variable)	UPMN	UPMN	UPMN	UPMN
Other Household Composition (Roster-Derived) Variables	UPMN	UPMN	UPMN	UPMN
Pair Relationship Variables and Multiplicity/Household Counts	PMN[10]	PMN	PMN	PMN

1 The 1999 survey year also included a paper-and-pencil interviewing (PAPI) sample. The procedures listed here are from the computer-assisted interviewing (CAI) sample.

2 "Random" refers to a random assignment within quarter for interview date, and a random assignment using age and interview date for birth date.

3 "None" means that no missing values were encountered after editing, so that no imputation was necessary. For gender (from the 2002 NSDUH onward) and age, missing values were precluded by design (see Chapter 4).

4 "UHD" refers to the unweighted sequential hot-deck method of item imputation described in this report (see Appendix A).

5 "MPMN" refers to the multivariate predictive mean neighborhood model-based procedure described in this report (see Appendix C).

6 "UPMN" refers to the univariate predictive mean neighborhood model-based procedure described in this report (see Appendix C).

7 "WHD" refers to the weighted sequential hot-deck method of item imputation described in this report (see Appendix A).

8 Although MPMN was the method used for health insurance in all years since the 1999 survey, the variables on which the imputation was applied changed in the 2002 NSDUH.

9 "Binge drinking" was defined as having five or more drinks on the same occasion on a given day.

10 "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 PMN Imputation Procedure for the NSDUH Sample

PMN was developed specifically for the 1999 survey. A combination of model-assisted imputation and a random nearest neighbor hot deck, PMN was implemented for nearly all variables requiring imputation in the 2002 NSDUH (exceptions are given in Exhibit 3.1).

In general, when large nonresponse occurs, limited donor sets can be used for imputation. For the 2002 NSDUH, to adjust for this sparseness of data, predictive mean modeling was used for the imputation of many of the variables (Exhibit 3.1). The models incorporated sampling design weights[4] with a response propensity adjustment computed to make the item respondent weights representative of the entire sample. The item response propensity model is a special case of the generalized exponential model (GEM), 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 gave a summary of the effects of covariates on the outcome variable. Unlike the sequential hot-deck imputation method, where values for the covariates were matched through a sorting procedure, the model-based approach used the predictive mean to convert the covariates' effects into a single number. The predictive means, along with other constraints, were used to define the neighborhoods from which donors were randomly selected for the final assignment of imputed values. This assignment either was done one value at a time (univariate predictive mean neighborhoods, or UPMN) or used several response variables at once (multivariate predictive mean neighborhoods, or 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 could not have been loosened. The latter, called "likeness constraints," could have been loosened if insufficient donors were available to meet the restriction. If more than one likeness constraint was placed on a neighborhood, the restrictions were loosened in a priority order deemed appropriate for the response variable in question.

In the 2002 NSDUH, because the drug use variables, as well as variables related to income, insurance, and household composition, were 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 use 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

[4] In the 2002 NSDUH, 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 demographic and drug imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.

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 have been considered a likeness constraint. However, this restriction was never loosened because the models were also built separately for the age groups. The imputation of missing values in the demographic variables was also performed within separate age groups: 12 to 17 year olds, 18 to 25 year olds, and persons 26 years of age or older. This was not due to a high correlation with age, but rather to the need to facilitate processing by decreasing the size of the datasets.

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 2002 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 in the hot-deck steps. The three State rank categories were defined in terms of the income level of the States: high-income States, middle-income States, and low-income States.

3.3 Other Imputation Procedures Used in the 2002 NSDUH

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

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

In the 2002 survey, for the first time, missing values were imputed in variables concerning immigrants. Respondents were asked whether they were born in the United States or not. Those respondents who were born outside the 50 States were also asked how long they had lived in the United States. Using this information, missing values were imputed in the indicator variable regarding whether the respondent was born in the United States, and in a derived variable giving the age of entry into the United States. A weighted hot-deck method (described in Appendix A), with weights unadjusted for missing values in these variables, was used to impute the missing values.

3.4 Changes in Procedures from the 2001 NHSDA to the 2002 NSDUH

Overall, the changes implemented between the 2001 survey and the 2002 survey were minor, both in number and in type. Some of these changes were the result of modifications to the CAI instrument. Others, however, were enhancements to procedures implemented in the 2002 survey, which were implemented as a result of a review of the procedures used in the 2001 survey. These enhancements involved both editing and imputation.

3.4.1 Differences Between Instruments in the 2001 NHSDA and the 2002 NSDUH Affecting Variables Requiring Imputation

For the first time in the 2002 survey, interviewers could no longer enter missing data for the gender question (QD01). Although the IRSEX variable name for gender was maintained for the sake of continuity, the IISEX imputation indicator was dropped.

Since the 2001 survey, the gender entry for the self in the questionnaire roster had to match the entry for gender given at the beginning of the questionnaire (QD01). In the 2002 instrument, interviewers had to also match the age entry for the self in the questionnaire roster with that in the non-roster portion of the questionnaire (denoted by the Blaise[5] variable CURNTAGE). It was possible to override this consistency check; however, the interviewer was required to explain why the consistency check was overridden. Other changes in the CAI logic were implemented in the questionnaire roster in the 2002 survey to improve internal consistency. These included a check requiring the respondent to have no more than one spouse (provided the spouses were not of different genders), to be younger than a parent or grandparent, and to be older than a child or grandchild. Each of these checks could have been overridden. The roster editing logic had to be adjusted to accommodate these new consistency checks. In most cases, a response which triggered a consistency check was changed by the interviewer to a more appropriate value. In the cases where the consistency checks were overridden, however, it was necessary to individually examine each explanation for an overridden consistency check, and to evaluate the legitimacy of the explanation. Depending upon the judgment of the legitimacy of the explanation, either an edit was applied or the data were left alone.

3.4.2 Improvements in Imputation Procedures from the 2001 Survey to the 2002 Survey

In nearly all the models, weights were adjusted for item nonresponse. These adjustments were determined by models called response propensity models, where age was often used as a covariate. Prior to the 2002 survey, the lifetime and recency-of-use response propensity models included age as a continuous covariate. To reduce problems associated with nonconvergence of these models, continuous age was replaced by categorical age variables with levels 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50+. In the models where age was left as a continuous variable, the

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

squared and cubed terms would also have been included. In the 2001 survey, most models that included continuous age as a covariate were "centered," where the mean age of respondents used to build a given model was subtracted from the given age, and the centered age was used as a covariate in the model. This alleviated multicollinearity problems, avoiding high standard errors and instability in the estimates. In the 2002 survey, centered age replaced the uncentered continuous age covariates in the remainder of models. More information on "centering" and "multicollinearity" can be obtained in Draper and Smith (1981, Section 5.5).

As indicated in Section 3.2, information about the State of residence of each respondent was incorporated in the modeling and hot-deck steps of the PMN procedure using three State usage-level categories for each drug, depending on the response variable of interest. In previous survey years, States were ranked according to their unweighted proportion of past month users (for the recency-of-use variables) and proportion of lifetime users (for the lifetime-usage variables). For the 2002 survey, States were ranked based on the weighted proportion for the characteristic of interest. In previous survey years, an additional problem occurred with rare drugs that caused the rankings to be unstable from year to year. In particular, many of the States had zero usage for the characteristic of interest, and different States would have this attribute from survey year to survey year. To alleviate this problem, State rank groups were determined using collapsed age groups for certain drug-age group combinations. In addition, if the number of States with zero usage for a characteristic of interest was more than a third of all States, all of the zero usage states were put into a single category, and the nonzero usage States were put into the other two categories.

Other minor improvements to the drug imputation programs for the 2002 survey included adding new logical constraints in the hot deck. These included a constraint that limited the number of possible imputed recency-of-use values when the respondent was interviewed on his/her birthday and had an age at first use that was one year less than his or her current age. Another new constraint was applied if the recipient did not use hallucinogens other than PCP, LSD, or Ecstasy. In this instance, the potential donor's hallucinogens age at first use must be equal to the minimum (reported or imputed) recipient's age of first use of PCP, LSD, and/or Ecstasy.

The most significant enhancement from the 2001 survey to the 2002 survey involved the methodology used to impute missing values in the health insurance variables. Prior to the 2002 survey, imputation was only applied to the overall health insurance and private health insurance variables. For the 2002 survey, however, imputations were performed on each of the constituent variables that were used to create the overall health insurance variable. The new overall health insurance variable, IRINSUR4, was therefore a recode of all the constituent imputation-revised health insurance variables. Details are available in Chapter 10.

3.4.3 Other Improvements in Procedures from the 2001 NHSDA to the 2002 NSDUH

A new feature that was implemented in its first phase for the 2002 survey involved the use of quality control checklists. These checklists were a more formal process to document quality control measures. Details regarding additional quality control measures, besides these formal checklists, are given in Appendix H. These quality control checklists incorporate all the steps required from first obtaining the necessary input variables to the final step of delivering the imputation variables. The checklists applied in the 2002 processing included checklists for modeling of drugs, income, and health insurance, and delivery of demographics, drugs, nicotine dependence, health insurance, and roster variables.

4. Core Demographics

4.1 Introduction

Several demographic characteristics were needed for all respondents in the 2002 National Survey on Drug Use and Health (NSDUH).[6] Core demographic data were collected on both the screener and the questionnaire. Missing values in screener and questionnaire demographic variables were imputed separately for the set of all eligible rostered individuals and for the set of completed respondents (i.e., screener data and questionnaire data were edited and imputed independently).[7] 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 2002 NSDUH discussed in this report are age, birth date, gender, race, Hispanicity, marital status, and education level (highest grade completed). The only noncore demographic variables imputed were the immigrant variables and employment status. Although the interview date was not classified as a core demographic variable, its editing procedures 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 were supplied, thus reducing the amount of statistical imputation required.[8] 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 interview date, age, and gender, no statistical imputation was required because no values were missing after editing. For birth date, 41 respondents had missing values, which were imputed using a random assignment from all possible birth dates that were consistent with the interview date and the age. The missing values in the marital status, race, Hispanicity, and education level variables were imputed using the predictive mean neighborhood (PMN) method. This procedure is described in greater detail in Appendix C. Missing values for the noncore employment status variables, which are discussed in the next chapter, were also imputed using the PMN method.

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

---

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

[7] See the weighting report for 2000 (Chen et al., 2002) for a description of the imputation procedures used for screener demographics for the set of all eligible rostered individuals. The procedures used for the 2000 survey and 2002 survey were equivalent.

[8] Logical editing undertaken to create base variables for imputation is described in this report; for more details on other editing performed on the 2002 NSDUH data prior to imputation, see Kroutil (2003a, 2003b, 2003c).

4.2 Variables Commonly Used as Covariates

In the PMN procedure, statistical modeling was performed to adjust weights for item nonresponse and also to calculate predictive means. The following variables were often used as covariates in both types of models for the PMN procedures. A complete list of covariates used in each model is available in Appendix E.

4.2.1 Household Type

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

4.2.2 Region

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

4.2.3 Segment ID

As described in the 2002 NSDUH Sample Design Report (Bowman et al., 2003), 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) was a two-letter State abbreviation followed by a two-digit FI region and a two-digit segment identifier, which uniquely identifies each segment. Although SEGID was not used as a covariate due to the large number of levels, it was used as a constraint in the hot-deck step of the PMN procedure for both race and Hispanicity, as noted in Section 4.4. For more information regarding segments, see the 2002 NSDUH: Sample Design Report.

4.2.4 Population Density

The population density variable PDEN2 was used to categorize segments according to modified 1990 census data, which was adjusted to more recent data from Claritas, Inc.[9] 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 Percentage 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.

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

4.2.6 Percentage 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 Percentage of Owner-Occupied Households

The owner-occupied household variable OWNOCONC was also used to categorize segments according to adjusted 1990 census data. It was used as a surrogate for income because wealthy segments tend to have many homeowners, while poor segments tend to have many renters. It 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 values was required to resolve inconsistencies and to fill in missing data. These edits are described below.

4.3.1 Edited Interview Date (INTDATE)

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

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

In some cases, the respondent's birthday occurred between the beginning and the end of the interview. In these cases, the interview date was set to the end-of-interview date stamp, which was consistent with the first date stamp after the respondent's birthday (this date stamp was indicated in the CAI).

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

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

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

Table 4.1 Interview Date Editing Summary

Value of EIIDATE	Assignment of Interview Date	Frequency	Percent
1	Begin date stamp (all date stamps exist)	68,071	99.92
1.01	Begin date stamp (all date stamps exist except last one)	8	0.01
1.02	Begin date stamp (all date stamps exist up through sedatives)	27	0.04
1.05	Begin date stamp (all date stamps exist up through pain relievers)	1	0.00
1.06	Begin date stamp (all date stamps exist up through inhalants)	1	0.00
2	Last existing date stamp (earlier than begin date stamp)	1	0.00
3	Tutorial date stamp (begin date stamp outside quarter)	1	0.00
6	Date later manually entered from RTI investigation	11	0.02
8	End date stamp (tutorial date stamp first after birthday)	1	0.00
8.07	End date stamp (marijuana 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 (noncore demographics date stamp first after birthday)	1	0.00
9	Begin date stamp (end date stamp in previous quarter)	1	0.00

4.3.2 Age

4.3.2.1 Final Edited Continuous Age (AGE)

After a respondent had entered his or her birth date in the first part of the questionnaire, he or she had multiple opportunities to change his or her age in response to consistency checks throughout the questionnaire. Therefore, it was possible for the age recorded by the respondent at the beginning of the questionnaire (CALCAGE) to be different from the age at the end of the questionnaire (NEWAGE). The final age variable, AGE, was determined using these two variables, in addition to three other sources: the age calculated from the 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 was given to CALCAGE, NEWAGE, and the age calculated from AGE1 and INTDATE. If the final age (AGE) did not agree with the originally entered birth date (AGE1), the birth date was also 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 were not equal, but NEWAGE was exactly equal to CALCAGE (adjusted by Blaise[10] to a changed interview date if the interview date was changed within the questionnaire), and the respondent's birthday did not fall between the dates corresponding to TBEG_TUT and INTDATE (age indicator = 1); else

NEWAGE, if nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, the given value of CALCAGE agreed with what it should be based on INTDATE and the given birth date (i.e., EIIDATE not equal to 6), and NEWAGE and CALCAGE were exactly equal (age indicator = 1); else

age calculated from INTDATE and the reported birth date, if the birth date was nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, and the given value of CALCAGE did not agree with what it should be based on INTDATE and the given birth date (EIIDATE = 6), where the newly calculated age based on INTDATE was exactly equal to the screener age and/or the roster age (if it existed) (age indicator = 2); else

NEWAGE, if NEWAGE differed from CALCAGE and NEWAGE = screener age and NEWAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 3); else

CALCAGE, if CALCAGE differed from NEWAGE and CALCAGE = screener age and CALCAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 4); else

age calculated from reported birth date and INTDATE, if EIIDATE = 5 and NEWAGE = CALCAGE (but neither was equal to the correct age) (age indicator = 5); else

NEWAGE, if NEWAGE differed from CALCAGE, but NEWAGE = roster age, provided roster age existed (age indicator = 6); else

CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = roster age, provided roster age existed (age indicator = 7); else

---

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

NEWAGE, if NEWAGE differed from age calculated from reported birth date and INTDATE, but NEWAGE = CALCAGE, screener age, and roster age (if it existed) (age indicator = 8); else

CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = age calculated from INTDATE and the reported birth date, and CALCAGE was within 1 year of screener age and roster age (age indicator = 9).

For a summary of the editing to create AGE for the 2002 NSDUH, see Table 4.2. This information was recorded in the editing indicator variable EIAGE.

Table 4.2 Age Editing Summary

Value of EIAGE	Assignment of Age	Frequency	Percent
1	NEWAGE (consistent with CALCAGE and INTDATE - AGE1)	68,122	99.99
2	Age from INTDATE and AGE1 (consistent with screener age)	2	0.00
4	CALCAGE (consistent with screener age)	1	0.00
6	NEWAGE (consistent with roster age)	1	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 to 17, 18 to 25, 26 to 34, and 35+), CATAG2 with three levels (12 to 17, 18 to 25, and 26+), and CATAG3 with five levels (12 to 17, 18 to 25, 26 to 34, 35 to 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, a birth date was randomly selected from all possible birth dates, given the final age and interview date. Each date in this period (365 or 366 days, depending on whether the period includes February 29 in a leap year) had an equal probability of selection.

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

Table 4.3 Birth Date Editing Summary

Value of EIBDATE	Assignment of Birth Date	Frequency	Percent
1	Reported birth date	68,084	99.94
2	Reported birthday, year from AGE and INTDATE	1	0.00
3	Randomly assigned using AGE and INTDATE	41	0.06

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 imputation-revised demographic variables. However, the edited variables were not released to the analytic and public use files; only imputation-revised variables were released to these files.

4.4.1 Gender

For the first time in the 2002 NSDUH, it was mandatory that an interviewer enter the respondent's gender in QD01. As a result, it was not possible to have missing values for this question. To maintain continuity with previous years, the variable IRSEX was created in the 2002 survey year. However, it was not necessary to create an imputation indicator, since IRSEX and QD01 were exactly equivalent.

4.4.2 Race

In the 2002 questionnaire, three core questions (QD05, QD05ASIA, and QD06) focused on the respondent's race and two focused on the respondent's ethnicity[11] (QD03 and QD04). In keeping with guidelines from the Office of Management and Budget (OMB),[12] "Hispanic/Latino" was considered an ethnicity, not a race. However, when given the opportunity to enter a race when the given choices did not apply, many respondents entered "Hispanic" or some Hispanic group, resulting in a considerable amount of missing data for the race question. The final drug-use tables were cross-classified with a variable that combined race and ethnicity. Nevertheless, separate variables were initially created for race and ethnicity, and the race/ethnicity variables used in the tables were derived from these separate variables. This subsection and the next three subsequent subsections outline how race and ethnicity were edited and imputed in the 2002 NSDUH.

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

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

4.4.2.1 Edited Race (EDRACE)

Respondents were given the choice of six categories in QD05 (white, black/African American, American Indian/Alaska native, native Hawaiian, other Pacific Islander, Asian, or some other race), of which they could have chosen more than one. If the "other" category was chosen, the interviewer was directed to manually enter the alternative to the given categories, denoted as the "other-specify" (or "alpha-specify") response, which was coded to correspond either to existing categories or to require imputation. (Details of the procedures to assign codes to responses and apply them to existing categories are described in Appendix D.) If the respondent identified himself/herself as Asian, he or she was routed to QD05ASIA, where one or more of the Asian categories were selected (Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, or some other Asian group). As with QD05, interviewers could have manually entered the alternative to the choices given, which would have been coded either to the existing categories or to require imputation. The coding scheme was the same for the alpha-specify responses for QD05 and QD05ASIA. That is, even though the specific Asian categories appeared in an additional question, the answers to QD05ASIA were treated exactly as if they came from QD05.[13] If multiple categories were selected in either or both of QD05 and QD05ASIA, the respondent was directed to QD06, where the respondent was asked to identify the single race with which he or she identified most closely.

Sometimes, the alpha-specify responses did not add any information. This occurred either when the response was not considered a race (e.g., "Hispanic," "American"), or when the response echoed information already conveyed in responses to the existing race categories (e.g., the respondent selected the "black/African American" category, and the interviewer wrote in "African" in the alpha-specify response). The former were considered "nonrace" entries, and the latter were considered "redundant" entries. In either case, the alpha-specify response was discarded. This method is an improvement over the algorithm from the 2001 NHSDA and earlier surveys, which incorrectly classified several respondents with noninformative or redundant other-specify codes as being of more than one race.

When the responses to QD05, QD05ASIA, and QD06 were combined to determine the single race with which a given respondent identified, 13 answer categories resulted (white, black/African American, American Indian/Alaska native, native Hawaiian, other Pacific Islander, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, other Asian, or some other race). However, the final race variable IRRACE was a four-level nominal variable: American Indian or Alaska Native, Asian or Pacific Islander, black, and white.[14] For respondents with more than one race, the response from QD06 was used if it existed. In this instance, a single race was assigned using a priority rule, where the highest priority was given to black/African

[13] The exception to this rule was with the response "Indian." If "Indian" was indicated in the alpha-specify response to QD05, he or she was classified as an American Indian. However, if "Indian" was indicated in the alpha-specify response to QD05ASIA, he or she was classified as Asian Indian. Details are in Appendix D.

[14] 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.

American, followed in priority order by Asian/Pacific Islander, American Indian/Alaska native, and white.[15]

EDRACE, the base variable for imputing race, was created using the following rules, under three possible scenarios:

Scenario 1: If only one category was identified in QD05, and if Asian was selected, only one Asian category was chosen in QD05ASIA, EDRACE =

> the single race identified in QD05, if that single race was not "other"; else
>
> race recode from alpha-specify response(s)[16] when "other" or "other Asian" was the only race selected in QD05, if a valid recode was available;[17] else
>
> missing.

Scenario 2: If more than one race was chosen in response to QD05 or QD05ASIA, EDRACE =

> the race response in QD06, if it was not "other," "other Asian," or missing; else
>
> race recode from alpha-specify response if QD06 = "other" or "other Asian" and a valid recode was available; else
>
> race assigned from the multiple responses given to QD05, using the following "priority rule": black/African American, Asian, American Indian/Alaska native, and white.

Scenario 3: If no response was given to QD05 (and hence QD05ASIA), EDRACE =

> race recode from alpha-specify response to QD04 (Hispanic origin group), if a valid recode was available; else
>
> missing.

---

[15] 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 was black/African American; otherwise, if Asian was among the groups selected, the single race for the respondent was Asian, etc. Details are given in Appendix D.

[16] QD04 (Hispanic-origin group question, see Section 4.4.5), 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 were 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.

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

4.4.2.2 Edited Race, Finer Categories (NEWRACE)

NEWRACE was a 15-level edited race variable used as a base variable for the final finer race-categories variable IRNWRACE. It was created by combining information from QD05 and QD05ASIA, but not from QD06. The other-specify response to QD04 was also used, if it corresponded to a valid race category and there was no other-specify response from QD05 or QD05ASIA. If the respondent gave a single response to QD05 and (if applicable) QD05ASIA, this response was used as a level in NEWRACE. This included 5 categories from QD05 (white, black/African American, American Indian/Alaska native, native Hawaiian, and other Pacific Islander), 7 categories from QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, and other Asian, provided the alpha-specify response to QD05ASIA was indeed an Asian group), and 3 categories representing combinations of the above 12 responses, 1 each for "native Hawaiian and other Pacific Islander," "Asian multiple category," and "more than one race," where the latter category did not include respondents who either were both native Hawaiian and other Pacific Islander, or were of multiple Asian races. The levels of NEWRACE were provided by the combined categories of QD05 and QD05ASIA, and three multiple race categories, as shown in Table 4.4.

NEWRACE was created in the following manner:

NEWRACE =

> 1-5, 7-13, if either this race category was the only one selected in QD05, or "other" and/or "other Asian" was the only race selected in QD05 and the alpha-specify response(s) was recoded to this race category, or QD05 was missing and the alpha-specify response from QD04 was recoded to this (single) race category;[18] else
>
> race assigned based on the census of a multiracial country of origin as stated in other-specify for QD05, provided "other" was the only race selected in QD05 and the country of origin was not Hispanic, where a random number was used to allocate a race; else
>
> 6, if either two selections, native Hawaiian and other Pacific Islander, were made in QD05, or the only race selected in QD05 was "other," with an alpha-specify response that was interpreted to be a combination of native Hawaiian and other Pacific Islander, or QD05 was missing and the alpha-specify response from QD04 was interpreted as a combination of native Hawaiian and other Pacific Islander; else
>
> 14, if either more than one race was selected in QD05 where all those selected were 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, or QD05 was missing and the alpha-

[18] An example where this could occur: if a respondent marked QD03 = 1 (Hispanic), but in the other-specify response to QD04 indicated "Haitian" as the Hispanic group, and did not answer QD05, he or she would have "black" as a race.

Table 4.4 Levels of NEWRACE

1	White
2	Black/African American
3	Native American or Alaska Native
4	Native Hawaiian
5	Other Pacific Islander
6	Native Hawaiian and other Pacific Islander
7	Chinese
8	Filipino
9	Japanese
10	Asian Indian
11	Korean
12	Vietnamese
13	Other Asian
14	Asian multiple category
15	More than one race

specify response from QD04 was interpreted as a combination of several Asian categories; else

15 (more than one race), if either 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; 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 two or more races; or QD05 was missing and the alpha-specify response from QD04 was interpreted as a combination of two or more races; else

missing.

Those respondents who indicated "Asian" in an other-specify response for race, but not one of the specific Asian groups, were assigned a code indicating that a finer Asian category needed to be imputed. This included respondents who indicated a country of origin, and were randomly allocated to "Asian." (These respondents would be included under "missing" above.)

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

The imputation-revised race variables were created using a multivariate predictive mean neighborhood (MPMN) method for imputation of missing values. The method was enhanced in the 2002 NSDUH, when conditional probabilities were used for the first time for a few item nonrespondents. The MPMN method as applied to the race variables is explained in detail in the next four subsections: setup for model building, computation of predictive means, assignment of imputed values, and constraints on MPMNs.

4.4.2.3.1 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. The separate age groups were used for ease of processing and consistency with other variables and not because of any strong correlation between age and race. Because all interview respondents were asked the race questions, no subsetting of the data was necessary.

Before predictive mean modeling was implemented, weights were adjusted for item nonresponse to the race questions. (In the 2002 NSDUH, the final analysis weights were used if they were available. However, because the final weight adjustments were not completed at the time of the demographic imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.[19]) An interview respondent was considered an item nonrespondent for race if either EDRACE was missing, NEWRACE was missing, or both. The weights of the item nonrespondents were redistributed among the item respondents using an item response propensity model. The item response propensity model is a special case of the generalized exponential model (GEM),[20] which is described in greater detail in Appendix B. A single response propensity model was used for all three age groups. The covariates in this model included census region, household type, final edited age (mean-centered), percentage Hispanic population, percentage non-Hispanic black population, and percentage of owner-occupied households.[21]

4.4.2.3.2 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.[22] The predictors included in the models were the same as those used in the item response propensity model for race.

For the youngest age group, the household type covariate was collapsed from three levels to two: Hispanics and non-Hispanic blacks were collapsed into a single category; the remaining category was non-Hispanic whites. For the oldest age group, three of the covariates—household type, percentage Hispanic population in the segment, and percentage non-Hispanic black population in the segment—were collapsed from three levels to two. In each case, the first two levels of the covariate were collapsed into one. (The household type variable was collapsed the same way as for the youngest age group. The combined categories for the other two variables were 0 to 50 percent and over 50 percent for non-Hispanic blacks, and 0 to 70 percent and over 70 percent for Hispanics.) The collapsing was done in order to stabilize the regression models,

---

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

[20] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

[21] Although a single response propensity model was used across all three age groups, separate response models were fitted within the three age groups. Because age was included as a covariate, the weights were still appropriately adjusted with a single response propensity model.

[22] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in the *SUDAAN® User's Manual, Release 8.0* (RTI, 2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

thus producing more reliable predictive means. The instability was caused by empty, or nearly empty, cells in a frequency table of each covariate by the response variable.

For example, the race of the householder (household type) was frequently equal to the race of the respondent. This was especially true for the oldest age group because the respondent and the householder were often the same person. As a result, when the race of the householder was not the same as the race of the respondent, empty or nearly empty cells occurred in the frequency table for some combinations of variables. By collapsing levels of the covariate, cells with low numbers were collapsed with other cells, reducing the imbalance.

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 (American Indian/Alaska native, Asian/Pacific Islander, black/African American). The probability of falling into the fourth race category (white) was not included because it was completely defined by the first three elements in the predictive mean vector being calculated as one minus their sum. A predictive mean vector of three predicted means was created from the polytomous logistic regression model.

The number of respondents who entered more than one race was very small compared to the other categories. For this reason, the models for race did not account for multiple race respondents as a separate category; rather, all respondents were assigned a single race as given in EDRACE. As stated earlier, for most respondents who identified more than one race, this single race was given by QD06, and the priority rule was used to determine the single race for the remainder. A handful of respondents were classified as "more than one race" based on an alpha-specify response, but no individual races were given, and no single race was given in QD06. (QD06 was skipped since the interviewer entered a single category corresponding to the alpha-specify response.) Since the priority rule was applied for other respondents who did not answer QD06, the priority rule was applied (in reverse) for these respondents as well. Using the priority rule, the single race for these respondents could not be white, since white had the lowest position in the priority order. (Respondents who entered "brown" were also in this category.) The assumption that these respondents would not consider themselves white was used to construct conditional probabilities. Instead of the usual three predicted means using the model's predicted probabilities directly, two predicted means were derived using conditional probabilities: the probability that the recipient was in the American Indian/Alaska native race category given that he/she was not white; and the probability that the recipient was in the Asian/Pacific Islander category given that he/she was not white.

4.4.2.3.3 Assignment of Imputed Values

For the race questions, the PMN method required the selection of an item respondent who was similar to each item nonrespondent. Specifically, the item respondent "donated" his or her value for EDRACE to the item nonrespondent. Most often, the selected item respondent, called the "donor," was randomly chosen from a "neighborhood" of potential donors. The item respondents in this neighborhood were the ones deemed to be most similar to the given item nonrespondent, who was called the "recipient." Item respondents who were deemed dissimilar to the recipient were discarded from the neighborhood by means of constraints. The predictive means calculated in the previous step were usually considered in these constraints. Because multiple variables were considered in the distance measure, "similarity" was defined in terms of the smallest Mahalanobis distance.[23] The PMN methodology is described in more detail in Appendix C; the constraints used for the race variables are described in the next section.

Separate assignments were performed within each of the three age groups. This type of age group-specific assignment was executed for almost all imputation-revised variables in the 2002 NSDUH. If the recipient had missing values for both EDRACE and NEWRACE, the donor gave values for both variables to the recipient. This ensured consistency between IRRACE and IRNWRACE.

4.4.2.3.4 Constraints on MPMNs

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

- If the recipient was of Hispanic origin, the donor must also have been of Hispanic origin.

- If the recipient was a member of a particular Hispanic origin group (e.g., Mexican, Puerto Rican, Central or South American, Cuban), the donor must also have been a member of that group. If the recipient was a member of more than one Hispanic origin group, the donor must have been a member of at least one of those specified by the respondent.

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

- If the recipient was known to be of more than one race, but the specific races were unknown and QD06 was not answered, it was assumed that the respondent would not answer "white" to the single race question. This was recorded in EDRACE. (This was an application of the priority rule in reverse: "white" had the lowest priority in the priority rule, so multiple-race respondents could not have been white in EDRACE.) Recipients with other-specify answers such as "brown" were also included in this category.

- In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that each of the donor's three predictive means (two when the recipients were multiple race or known to be nonwhite), as described in Section 4.4.2.3.2, must have been within 5 percent (within "delta") of each of the recipient's three predictive means. If no potential donors met both of the above conditions for a particular item nonrespondent, the constraint on the segment of

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

the potential donor was removed first. If no potential donors met the "delta constraint," the delta constraint was also removed. 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 F.

4.4.2.4 Imputation and Editing Summary for Race

To differentiate the final imputed values from nonmissing values, a concomitant indicator variable, II2RACE, indicated how the levels of IRRACE were derived. II2RACE was a more detailed version of the variable IIRACE, which was the only imputation indicator variable for IRRACE available for the 1999 NHSDA. Table 4.5 gives the levels for both IIRACE and II2RACE, and shows how the levels of II2RACE mapped to those of IIRACE. The 15-level race variable, IRNWRACE, also had a concomitant indicator variable. Table 4.6 summarizes the levels of IINWRACE, the concomitant indicator variable for IRNWRACE.

Table 4.5 IRRACE Editing and Imputation Summary

Value of II2RACE	Assignment of IRRACE	Frequency	Percent	Level of IIRACE
1	From single QD05 response	64,202	94.24	1
2	From QD06 response	1,530	2.25	1
3	Logically assigned from alpha-specify response	342	0.50	2
4	Assigned with census data from country of origin	201	0.30	3
5	Single race determined from multiple responses	116	0.17	1
6	Statistically imputed (unrestricted)	50	0.07	4
7	Statistically imputed (restricted to Hispanic groups)	1,685	2.47	5

Table 4.6 IRNWRACE Editing and Imputation Summary

Value of IINWRACE	Assignment of IRNWRACE	Frequency	Percent
1	From QD05 response(s)	65,847	96.65
2	Logically assigned from alpha-specify response(s)	348	0.51
3	Assigned with census data from country of origin	196	0.29
4	Statistical imputation of "Asian" into finer categories	17	0.02
5	Statistically imputed (unrestricted)	43	0.06
6	Statistically imputed (restricted using Hispanicity)	1,675	2.46

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 described 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 was not "other"; else

Hispanic-origin group recode from alpha-specify response(s),[24] if "other" was selected and a valid recode was available;[25] 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 was available; else

missing.

The base variable for creating an imputation-revised Hispanic-origin indicator was EDHOIND, which was created using responses to QD03 and the edited Hispanic-origin group variable (EDQD04) as follows:

EDHOIND = 1 (Hispanic), if QD03 = 1 OR if alpha-specify response to QD05 indicated that the respondent was Hispanic OR if EDQD04 had a value indicating that the respondent was Hispanic; else

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

---

[24] Both QD04 (Hispanic-origin group question) and QD05/QD06 allowed respondents to specify a race or Hispanic-origin group, respectively, other than those listed in the questions, when they selected 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.

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

missing.

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 predictive means, 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. The separate age groups were used more for ease of processing and consistency with other variables rather than due to any strong correlation between age and Hispanic origin. Because all interview respondents were asked the question about Hispanic origin, no subsetting of the data was necessary.

As for the race variables, weights were adjusted for item nonresponse to the Hispanic origin question, QD03, using an item response propensity model. (For these race variables, weights were defined in a similar manner to the way weights were determined for other demographic variables. Details on how the weights were defined can be found in Section 4.4.2.3.1.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model were census region, imputation-revised race, percentage Hispanic population, percentage non-Hispanic black population, and percentage of owner-occupied households. As with race, a single item response propensity model was used across all age groups.

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 (mean-centered), age squared, age cubed, percentage Hispanic population, percentage non-Hispanic black population, and percentage 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 stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated 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 F for the numbers of respondents that met 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 4.7 summarizes item nonresponse for the Hispanic-origin indicator. This information was recorded in the variable IIHOIND.

Table 4.7 Hispanic-Origin Indicator Editing and Imputation Summary

Value of IIHOIND	Assignment of IRHOIND	Frequency	Percent
1	From questionnaire	67,942	99.73
2	From alpha-specify responses	59	0.09
3	Statistically imputed	125	0.18

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 that were similar to their counterparts from years prior to the 1999 NHSDA: 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).

Additional recodes that used IRHOIND incorporated information about respondents who indicated membership in more than one race. The variable NEWRACE1 was similar to the detailed race variable IRNWRACE, except that Hispanic respondents were separated out and given their own level. Detailed race information in NEWRACE1was therefore only available for non-Hispanic respondents. In particular,

NEWRACE1 = IRNWRACE, if IRHOIND = 2; else

16 (Hispanic), if IRHOIND = 1.

Three other variables were derived from NEWRACE1. These variables were EXPRACE, NEWRACE2, and RACE4. EXPRACE was created by collapsing the categories that could have contained respondents of different races (Hispanic, all multiple category levels, Hawaiian/other Pacific Islander, and other Asian). In particular, EXPRACE

=	1	Non-Hispanic white (NEWRACE1 = 1)
=	2	Non-Hispanic black (NEWRACE1 = 2)
=	3	Non-Hispanic American Indian/Alaska Native (NEWRACE1 = 3)
=	4	Non-Hispanic Native Hawaiian (NEWRACE1 = 4)
=	5	Non-Hispanic Other Pacific Islander (NEWRACE1 = 5)
=	6	Non-Hispanic Chinese (NEWRACE1 = 7)
=	7	Non-Hispanic Filipino (NEWRACE1 = 8)
=	8	Non-Hispanic Japanese (NEWRACE1 = 9)
=	9	Non-Hispanic Asian Indian (NEWRACE1 = 10)
=	10	Non-Hispanic Korean (NEWRACE1 = 11)
=	11	Non-Hispanic Vietnamese (NEWRACE1 = 12)
=	12	Other (NEWRACE1 = 6, 13, 14, 15, 16)

By collapsing all the Asian categories in NEWRACE1 into a single category, and collapsing the native Hawaiian and other Pacific Islander categories, the following levels were derived in the variable NEWRACE2.

NEWRACE2

=	1	Non-Hispanic white (NEWRACE1 = 1)
=	2	Non-Hispanic black (NEWRACE1 = 2)
=	3	Non-Hispanic American Indian/Alaska Native (NEWRACE1 = 3)
=	4	Non-Hispanic Pacific Islander (NEWRACE1 = 4, 5, 6)
=	5	Non-Hispanic Asian (NEWRACE1 = 7-14)
=	6	Non-Hispanic more than one race (NEWRACE1 = 15)
=	7	Hispanic (NEWRACE1 = 16)

Finally, the variable RACE4 was very similar to the recoded variable RACE, except that it used NEWRACE1 rather than IRRACE and IRHOIND. The only visible difference with RACE transpired when people of more than one race were allocated a race based upon their response to QD06, or when the priority rule was used. In RACE4, respondents of more than one race were placed in the "other" category.

RACE4

=	1	Non-Hispanic white, single race (NEWRACE1 = 1)
=	2	Non-Hispanic black, single race (NEWRACE1 = 2)
=	3	Hispanic (NEWRACE1 = 16)
=	4	Non-Hispanic other or more than one race (all other values of NEWRACE1)

4.4.5 Hispanic-Origin Group

4.4.5.1 Edited Hispanic-Origin Group (EDHOGRP and EDHOGRP2)

The Hispanic-origin group variables divided respondents of Hispanic origin into finer categories. Two edited Hispanic-origin group variables were created: one for the purposes of modeling, and the other for the purposes of the final assignment of imputed values. For the final assignment of imputed values, all information from EDQD04 was retained, so that EDHOGRP and EDQD04 were virtually equivalent.[26] However, the model that was used to determine the assignment of imputed values required collapsing of levels. Hence, a new variable, EDHOGRP2, was created to act as the response variable. In the model for Hispanic origin, all Hispanics who were not from Puerto Rico, Mexico, or Cuba were collapsed into a single group. Hispanic respondents for whom the Hispanic group was unknown, but for whom partial information was available, could not have been included in the model because they were still considered as item nonrespondents. Hence, EDHOGRP2 included levels for Puerto Rico, Mexico, Cuba, all other Hispanics, and a level indicating that the respondent was an item nonrespondent.

4.4.5.2 Imputation-Revised Hispanic-Origin Group (IRHOGRP3)

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. The predictive mean vector had only three elements associated with the first three levels of EDHOGRP2: 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 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.

4.4.5.2.1 Setup for Model Building

All respondents with IRHOIND = 2 were automatically assigned IRHOGRP3 = 99 and were excluded from the item response propensity model, the predictive mean model, and the set of potential donors. In contrast to the other demographic variables, imputations were not conducted separately within the three age groups. This was done for two reasons. First, for the 2002 NSDUH, there was a weak statistical relationship between the Hispanic-origin group and age. Therefore, separating the respondents into age groups was unlikely to improve the probability that a donor who was similar to the item nonrespondent would have been found. Second, only respondents with IRHOIND=1 were eligible to be donors, so keeping all age groups in the same data set keeps the size of the donor pool large enough to ensure that a suitable donor would have been found.

An interview respondent was considered an item nonrespondent for Hispanic-origin group if his or her value for EDHOGRP2 was missing. The weights of the item nonrespondents

---

[26] Differences were limited to the ordering of levels, and the level assigned to "no information available" was set to 10 in EDQD04 and to missing in EDHOGRP.

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 (mean-centered), age squared, age cubed, percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, the interaction of age and gender, and the interaction of age squared and gender.

4.4.5.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each of the first three Hispanic-origin group categories was modeled for all age groups together, using polytomous logistic regression. The predictors included in the model were census region, imputation-revised race, gender, age, age squared, age cubed, percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, the interaction of age and gender, and the interaction of age squared and gender.

4.4.5.2.3 Assignment of Imputed Values

All age groups were aggregated in this step, for the reasons given in Section 4.4.5.2.1. The constraints used to select donors are described in the next section.

4.4.5.2.4 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 or 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. This was possible by using the variable EDHOGRP instead of EDHOGRP2, where the race information was embedded in the levels of the variable.

In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that each of the donor's three predictive means, as described in Section 4.4.5.2.2, 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. If still no donor could be found, the constraint on the predictive means was also removed. See Appendix F for the numbers of respondents that met each set of likeness constraints on sets of eligible donors.

4.4.5.3 Imputation and Editing Summary for Hispanic-Origin Group

To differentiate the final imputed values from nonmissing values, a concomitant indicator variable, II2HOGR3, gave the source of information for IRRACE. The levels of II2HOGR3 are summarized in Table 4.8. As was the case with IIRACE and II2RACE, a variable that gave somewhat less information, IIHOGRP3, was created for the 1999 NHSDA to give the source of information for IRHOGRP3. For the sake of consistency, this variable was again created for the 2002 NSDUH. Table 4.8 shows how the levels of II2HOGR3 mapped to those of IIHOGRP3. As with IRRACE, a priority rule[27] was used to determine what group a respondent belonged to if he or she gave more than one response. II2HOGR3 recorded these cases, whereas IIHOGRP3 merely considered these cases as a "response from questionnaire."

Table 4.8 Hispanic-Origin Group Editing and Imputation Summary

Value of II2HOGR3	Assignment of IRHOGRP3	Frequency	Percent	Level of IIHOGRP3
1	From questionnaire	8,005	11.75	1
2	From alpha-specify response(s)	642	0.94	2
3	Single Hispanic group determined from multiple responses	93	0.14	1
4	Statistically imputed (unrestricted)	60	0.09	3
5	Statistically imputed (restricted by IRRACE)	11	0.02	4
9	Legitimate skip (respondent was not Hispanic)	59,315	87.07	9

4.4.5.4 Hispanic-Origin Group Recodes

HISPGRP 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).

4.4.6 Marital Status

4.4.6.1 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 was 15 years old or older; else

99 (legitimate skip) if the respondent was younger than 15; else

missing.

4.4.6.2 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 was a four-level nominal variable. The four substantive levels of the imputation-revised marital status variable, IRMARIT, were the same as the four answer categories 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.

[27] The priority rule was the same as that used in past NSDUHs: Mexican, Cuban, Puerto Rican, Central/South American, Caribbean Islander, and other Hispanic. Details are given in Appendix D.

The full predictive mean vector had three elements corresponding to QD07 = 1, QD07 = 2 or 3, and QD07 = 4. The main differences between marital status imputation and race imputation were the relative simplicity of the editing process (Kroutil, 2003a, 2003b, 2003c) 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.

4.4.6.2.1 Setup for Model Building

Imputations were conducted separately within the same three age groups as for most of the other demographic variables. All respondents with AGE younger than 15 were assigned IRMARIT = 99. 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 or her value for EDMARIT was missing. The weights of the item nonrespondents 15 or older were reallocated to the item respondents 15 or older, using an item response propensity model. (Weights were defined in the same way as with other demographic variables. See the discussion about how the weights were defined in Section 4.4.2.3.1.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, population density, age (mean-centered), percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, and the interaction of age and gender.

4.4.6.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each marital status category was modeled for all age groups together using polytomous logistic regression.[28] The predictors included in the model were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, population density, age (mean-centered), percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, and the interaction of age 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

No logical constraints were used in defining neighborhoods for the marital status variable; only likeness constraints were utilized. In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first constraint required each

[28] All age groups were modeled together because the distributions of the answers for the youngest two age groups were lopsided, making it difficult to find convergent models.

of the donor's three predictive means, as described in Section 4.4.6.2.2, 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. The second constraint required donors and recipients to have an age difference of three years or less. See Appendix F 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 Marital Status

See Table 4.9 for a summary of item nonresponse for marital status (recorded in the variable IIMARIT).

Table 4.9 Marital Status Editing and Imputation Summary

Value of IRMARIT	Assignment of Marital Status	Frequency	Percent
1	From questionnaire	55,944	82.12
3	Statistically imputed	14	0.02
9	Legitimate skip (≤14 years old)	12,168	17.86

4.4.6.4 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).

4.4.7 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)

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

4.4.7.2.1 Setup for Model Building

The imputations for the highest grade completed variable 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 this question, no subsetting of the data was necessary.

Weights were adjusted for item nonresponse to the highest grade completed question, QD11. The covariates in the item response propensity model (see Appendix B for the more general GEM) were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, age (mean-centered), age squared, the interaction of age and gender, the interaction of age squared and gender, percentage Hispanic population, percentage non-Hispanic black population, and percentage of owner-occupied households.

4.4.7.2.2 Computation of Predictive Means

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

Using the adjusted weights, the probability of the respondent falling into each level of the response variables was modeled using polytomous logistic regression. The respondents aged 12 to 17 years old were modeled separately from the two older age groups. For the youngest age group, the predictors included in the model were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, percentage Hispanic population, percentage non-Hispanic black population, and percentage of owner-occupied households. For the other two age groups, the predictors included in the model were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, age (mean-centered), age squared, age cubed, the interaction of age and gender, the interaction of age squared and gender, percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, and imputation-revised marital status.

4.4.7.2.3 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.

4.4.7.2.4 Constraints on MPMNs

One logical constraint was used in defining neighborhoods: If the recipient was 12 to 25 years old, the donor must be the same age as the recipient. In the first attempt to find a neighborhood for each item nonrespondent, two likeness constraints were used. The first likeness constraint stated that the donor must have lived in the same segment as the recipient. The second likeness constraint stated that the donor's predictive means, as described in Section 4.4.7.2.2, must have been within 5 percent of the recipient's 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 potential donors could still not be found, the delta constraints were removed. See Appendix F for the numbers of respondents meeting each set of likeness constraints on sets of eligible donors.

4.4.7.3 Imputation and Editing Summary for Highest Grade Completed

Table 4.10 summarizes item nonresponse for the highest grade completed variable. This information was recorded in the variable IIEDUC.

Table 4.10 Highest Grade Completed Editing and Imputation Summary

Value of IIEDUC	Assignment of IREDUC	Frequency	Percent
1	From questionnaire	68,113	99.98
3	Statistically imputed	13	0.02

4.4.7.4 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

For the 2002 National Survey on Drug Use and Health (NSDUH),[29] missing values were imputed in two sets of variables in the noncore demographics module: the immigrant status and employment status variables. Additionally, the core demographics that were imputed in the 2002 NSDUH are discussed in Chapter 4.

For immigrant status, two variables, BORNINUS and ENTRYAGE, had missing values that were imputed. These variables recorded whether a respondent was born in the United States, and, if not, the age of entry into the United States. The imputation-revised versions of these variables were called IRBORNUS and IRENTAGE, respectively. The final goal was to create a data file containing variables that would have indicated whether a respondent could have been included in incidence analyses based on their immigrant status.

The variables describing current employment status were determined from multiple questions in the noncore demographics module. Instead of a single question asking the respondent to describe his or her "current" employment status, several questions were asked regarding the respondent's employment situation during the week preceding the interview and whether that week was atypical. The employment status questions were asked only of respondents aged 15 or older. A single imputation-revised variable, EMPSTATY, was created from the series of employment status questions. Unlike other imputation-revised variables, for historical reasons this variable was not preceded by an "IR" prefix. However, it was accompanied by an imputation indicator that did have the requisite "II" prefix, IIEMPSTY.

Respondents who either worked during the week preceding the interview or said they had a job were asked to write in the industry for which they worked, their occupation, and their main duties at work. Edited versions of the responses to some of these questions are discussed in a separate document (Kroutil, 2003a). Even though responses were edited, missing values were not imputed.

5.2 Immigrant Status

The edited immigrant status variables used to create IRBORNUS and IRENTAGE are described in Section 5.2.1. The edited variable BORNINUS, the base variable used for creating IRBORNUS, was derived from the questionnaire questions QD14 and QD15, and is described in Section 5.2.1.1. LIVEDUSA, and its derived continuous form, LENGTHLIV, were derived from questionnaire question QD16, and are used to create the base variable for IRENTAGE, ENTRYAGE. The variables LIVEDUSA and LENGTHLIV are discussed in Section 5.2.1.2, and ENTRYAGE is discussed in Section 5.2.1.3. Variables that were created specifically for the imputation of missing values in the immigrant variables are described in Section 5.2.2. In

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

Section 5.2.3, the methodology for imputing missing values in the variables BORNINUS (Section 5.2.3.1) and ENTRYAGE (Section 5.2.3.2) is discussed. The resulting variables are called IRBORNUS and IRENTAGE, respectively, and are used to create recoded variables for the purposes of analysis.

5.2.1 Edited Immigrant Status Variables

5.2.1.1 BORNINUS

All respondents were asked in QD14 whether they were born in the United States (excluding U.S. territories). Responses were limited to "yes" or "no," and if the response was no, the respondent was asked to name the country of origin in QD15. The edited variable BORNINUS was created using the responses to QD14. As part of the standard editing procedures, if the interviewer entered a U.S. State in QD15, the "no" in QD14 was overwritten with a logically assigned "yes." Other levels of BORNINUS were standard NSDUH missing data codes corresponding to "don't know," "refused," or "blank." More details about editing procedures are provided in a separate document (Kroutil, 2003a).

5.2.1.2 LIVEDUSA and LENGTHLIV

The following question (QD16) asks the length of time that the respondent has lived in the United States. Responses were given in ranges, which corresponded to categories provided by the following question:

About how long have you lived in the United States?

1 6 MONTHS OR LESS
2 MORE THAN 6 MONTHS BUT LESS THAN 1 YEAR
3 AT LEAST 1 YEAR BUT LESS THAN 5 YEARS
4 AT LEAST 5 YEARS BUT LESS THAN 10 YEARS
5 AT LEAST 10 YEARS BUT LESS THAN 15 YEARS
6 15 YEARS OR MORE

LIVEDUSA

The edited form of QD16 was given by the variable LIVEDUSA, which included these 6 categories, plus categories for missing values (identified by codes for "don't know," "refused," "blank," and "bad data"). A valid response was replaced by a bad data code if it was inconsistent with the respondent's age.

LENGTHLIV

In order to get a continuous estimate of how many years an immigrant had lived in the United States, each discrete category needed to be converted to a random number of years that was within the appropriate interval. While the lower bound for this interval could be obtained directly from the category selected, some derivation was required to determine the upper bound. This upper bound, denoted by "X," was created in the following manner:

- If LIVEDUSA=1 then X = 0.5

- If LIVEDUSA=2 then X = 1
- If LIVEDUSA=3 then X = 5
- If LIVEDUSA=4 then X = 10
- If LIVEDUSA=5 then X = min(15, CONTAGE-0.00274)
- If LIVEDUSA=6 then X = CONTAGE-0.00274 (where 0.00274 =1/365 since at a minimum the R has to be at least 1 day old when they entered the country)
- Else X = .

For those with BORNINUS = no (which was response 2 in QD14), the variable LENGTHLIV was created to randomly impute an actual length of time each immigrant had lived in the United States. The LENGTHLIV variable, based on the value of LIVEDUSA, was calculated using the following logic.

- If LIVEDUSA=1 then LENGTHLIV= 0+(X-0)*UNIF(0,1)
- If LIVEDUSA=2 then LENGTHLIV= .5+(X-.5)*UNIF(0,1)
- If LIVEDUSA=3 then LENGTHLIV= 1+(X-1)*UNIF(0,1)
- If LIVEDUSA=4 then LENGTHLIV= 5+(X-5)*UNIF(0,1)
- If LIVEDUSA=5 then LENGTHLIV= 10+(X-10)*UNIF(0,1)
- If LIVEDUSA=6 then LENGTHLIV= 15+(X-15)*UNIF(0,1)
- Else LENGTHLIV = .

UNIF(0,1) denotes a uniform random number generated between the values of 0 and 1.

5.2.1.3 ENTRYAGE

The variable ENTRYAGE represents the age that an immigrant entered the United States. The continuous age variable, CONTAGE, was defined as CONTAGE = (interview date – birth date + 1) / 365.25. This variable was created so that the variable ENTRYAGE would have had a continuous value, in the same manner that LENGTHLIV had a continuous value. ENTRYAGE was then calculated as ENTRYAGE = CONTAGE – LENGTHLIV.

5.2.2 Covariates Used in the Imputation of Immigrant Status Variables

Two variables were created specifically to aid in the imputation of missing values in the immigrant status variables.

HISPGRP2

A significant number of respondents who were born outside the United States were of Hispanic origin, with varying degrees of immigration depending upon the Hispanic group. A natural candidate for a classing variable, therefore, would have been a Hispanic group variable that had a separate level for "non-Hispanic." The variable IRHOGRP3 was the imputation-revised Hispanic group variable, the creation of which is described in Chapter 4. However, the levels of IRHOGRP3 were too fine to be used as imputation classes. As an alternative, a collapsed version of IRHOGRP3 was created, called HISPGRP2. It had four levels: 1 = Puerto Rican (IRHOGRP3 = 1), 2 = Mexican (IRHOGRP3 = 2), 3 = Other Hispanic (IRHOGRP3 = 3,4,5,6), 4 = Non-Hispanic (IRHOGRP3 = 99).

AGEADULT

The immigrant status of a respondent was also closely related to that respondent's age. Clearly, the age of entry of the immigrant into the United States was limited by the age of the respondent. The variable AGEADULT was created by collapsing AGE (the creation of which is described in Chapter 4), into five categories: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or over. AGEADULT is equivalent to the variable CATAG3, which is described in Section 4.3.2.2.

5.2.3 Imputation-Revised Immigrant Status Variables

5.2.3.1 IRBORNUS

The weighted hot-deck imputation procedure with a serpentine sort was used to impute a value of 1 or 2 for the 35 individuals who did not have valid interview responses for the BORNINUS variable. Potential donors were partitioned into imputation classes using HISPGRP2. For the serpentine sort, IRRACE (a four-level race variable, described in Chapter 4) and AGEADULT were used as sort variables. In the hot-deck procedure, possible donors were defined as BORNINUS in (1, 2), where donors in the imputation class were weighted using the sample design weights, appropriately adjusted for item nonresponse, extreme values, and calibrated to U.S. census control totals. IIBORNUS was an imputation indicator variable that contained a '1,' '2,' and '3' if IRBORNUS was questionnaire data, logically assigned data, and statistically imputed data, respectively. Table 5.1 gives a breakdown of the reported and imputed values in IRBORNUS.

Table 5.1 Reported and Imputed Values in IRBORNUS

Imputation-Revised BORNINUS		
IIBORNUS	**IRBORNUS**	**Frequency**
1	1	61,460
1	2	6,619
2	1	12
3	1	26
3	2	9
		68,126

5.2.3.2 IRENTAGE

As with IRBORNUS, the weighted hot-deck imputation procedure with a serpentine sort was used to impute an ENTRYAGE for the individuals with missing values. These included nine respondents who had missing values for IRBORNUS, but were imputed to have been born outside the United States, and three respondents who responded that they were born outside the United States, but did not have a valid age of entry in the country. Potential donors were partitioned into imputation classes using AGEADULT and the same Hispanic group variable as with BORNINUS, the HISPGRP2 variable. For the serpentine sort, IRRACE and AGE were used as sort variables. In the hot-deck procedure, possible donors were limited to immigrants, where the value for the variable ENTRYAGE was both positive and valid. The number of donors within each imputation class (HISPGRP2*AGEADULT) ranged from 27 to 1,122 individuals.

5.3 Current Employment Status

The edited employment status variables used to create EMPSTATY are described in Section 5.3.1. Section 5.3.1.1 discusses the edited variables JBSTATR and WRKHRSUS. Section 5.3.1.2 discusses the creation of EDEMPY, the base variable for imputation. Sections 5.3.2 and 5.3.3 discuss the imputation procedure for EMPSTATY, and Section 5.3.4 discusses the creation of EMPSTAT4, a recoded version of EMPSTATY.

5.3.1 Edited Employment Status Variables

5.3.1.1 JBSTATR and WRKHRSUS

The main edited variable used to summarize the respondent's current work situation was JBSTATR, which was subsequently used to create EMPSTATY. This edited variable combined information from QD26, QD29, QD30, QD31, QD32, and QD33. The categories for JBSTATR are shown in Table 5.2. WRKHRSUS was an edited variable created from QD29, which asks, "Do you **usually** work 35 hours or more per week at **all** jobs or businesses?" WRKHRSUS was used in some cases to determine whether employed respondents were employed full-time or part-time. Both variables are described in more detail in Kroutil (2003b).

Table 5.2 Categories of JBSTATR

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

[1] National Survey on Drug Use and Health

5.3.1.2 EDEMPY

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

EDEMPY =

99, if the respondent is 12 to 14 years old; else

1 (full-time), if JBSTATR = 1 or 190, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 1; else

2 (part time), if JBSTATR = 2 or 191, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 2; else

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

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

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

missing.

5.3.2 Imputation-Revised Employment Status (EMPSTATY)

Missing values in the edited employment status variable EDEMPY were replaced with imputed values using a multivariate predictive mean neighborhood (MPMN) procedure. This procedure is described in greater detail in Appendix C. The MPMN method was applied to employment status variables for the first time in the 2001 NHSDA; it was enhanced in the 2002 NSDUH to account for partial knowledge of employment status.

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

5.3.2.1 Setup for Model Building

Similar to the imputations that were performed on other demographic variables, imputations for employment status variables were conducted separately within the same three age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents aged 26 or older. All respondents with AGE younger than 15 were assigned EMPSTATY = 99. Only interview respondents with AGE of 15 or greater were considered as donors.

An interview respondent was considered an item nonrespondent for employment status if his or her value for EDEMPY was 5 (employed, part time versus full time unclear) or missing. The weights of the item nonrespondents 15 or older were reallocated to the item respondents 15 or older. (In the 2002 NSDUH, the final analysis weights were used if they were available.

However, because the final weight adjustments were not completed at the time of the demographic imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.[30]) The item response propensity model is a special case of the generalized exponential model (GEM),[31] which is described in greater detail in Appendix B. A single item response propensity model was used for all three age groups.[32] The covariates in the model were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, age (mean-centered), age squared, the interaction of age and gender, the interaction of age squared and gender, percentage Hispanic population, percentage non-Hispanic black population, and percentageof owner-occupied households.

5.3.2.2 Computation of Predictive Means

Using the adjusted weights, the probability of selecting each employment status category (employed full-time, employed part-time, unemployed, and other) was modeled using polytomous logistic regression.[33] Respondents aged 15 to 25 were modeled separately from respondents aged 26 or older.[34] The predictors included in the model for the respondents aged 15 to 25 were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, age (mean-centered), age squared, the interaction of age and gender, the interaction of age squared and gender, percentage Hispanic population, percentage non-Hispanic black population, and percentage of owner-occupied households. The predictors included in the model for the respondents aged 26 and older were census region, imputation-revised race, imputation-revised Hispanic-origin indicator, gender, age (mean-centered), the interaction of age and gender, percentage Hispanic population, percentage non-Hispanic black population, percentage of owner-occupied households, and imputation-revised marital status. The predictive mean vector used in the imputation procedure had three elements (three predictive probabilities) corresponding to the first three employment status categories.

5.3.2.3 Assignment of Imputed Values

The imputations were performed separately within each of three age groups: 15 to 17 year olds, 18 to 25 year olds, and respondents aged 26 or older. The relative ages of donors and recipients were also restricted based on a logical constraint. All constraints used to select donors are described in the next section.

---

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

[31] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

[32] Although a single response propensity model was used across all three age groups, separate predictive mean models were fitted within the three age groups. Because age was included as a covariate, the weights were still appropriately adjusted with a single response propensity model.

[33] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in the *SUDAAN® User's Manual, Release 8.0* (RTI, 2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

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

5.3.2.4 Constraints on MPMNs

Two logical constraints were used in defining neighborhoods for the employment status variable. These constraints were:

- The donor's age must be within 4 years of the recipient's age.
- If the recipient had EDEMPY = 5, the donor must have been employed either part-time or full-time (EDEMPY = 1 or 2).

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

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

5.3.3 Imputation and Editing Summary for Employment Status

See Table 5.3 for a summary of item nonresponse for employment status. For the first time in the 2002 NSDUH, a detailed imputation indicator (II2EMSTY) was created to separate the item nonrespondents into the two missingness patterns: those with EDEMPY missing and those with EDEMPY = 5. The table also shows how the levels of II2EMSTY map to those of IIEMPSTY.

Table 5.3 EMPSTATY Editing and Imputation Summary

Assignment of EMPSTATY	Frequency	Percent	Value of IIEMPSTY	Value of II2EMSTY
From questionnaire	55,906	82.06	1	1
Statistically imputed (unrestricted)	36	0.05	3	3
Statistically imputed (restricted to full time or part time)	16	0.02	3	4
Legitimate skip (respondent was 12-14 years old)	12,168	17.86	9	9

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

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

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

6. Drug Imputations

6.1 Introduction

Major changes were introduced in the imputation procedures for the drug use variables in the computer-assisted interviewing (CAI) sample of the 1999 National Household Survey on Drug Abuse (NHSDA).[35] In particular, a new imputation methodology (i.e., predictive mean neighborhood [PMN]) was developed specifically for the NSDUH. This methodology is a combination of weighted regression and nearest neighbor hot-deck imputation, where the hot deck is random whenever possible.[36] Its application to the drug use variables for the 2002 survey was similar to that of previous survey years, as is explained in the following sections.

This chapter describes how the PMN technique was applied to the drug use variables. In some cases, imputations were required because the respondent did not answer a given question. However, other responses were altered in the editing process due to inconsistencies. In these cases, the original response was 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 have been edited to past year usage, where past-month versus past-year-but-not-past-month use could have been determined by imputation. These editing processes are summarized by Kroutil (2003a).

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 predictive 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 predictive mean, the donor with the closest predictive mean was chosen.) The neighborhoods were created based on a single predictive mean (a univariate predictive mean neighborhood [UPMN]), or using several predictive means at once (a multivariate predictive mean neighborhood [MPMN]). Even if the neighborhood was constructed from a univariate predictive mean, the assignment of imputed values could have been either univariate or multivariate. The members of the neighborhood were restricted to satisfy two types of constraints: "logical constraints" and "likeness constraints." Constraints that made the imputed values consistent with preexisting values of other variables were called logical constraints and were required for the candidate donor to be a member of the neighborhood. Likeness constraints were implemented to make donors and recipients as much alike as possible. Although logical constraints could not have been loosened, likeness constraints could have been loosened if they forced the donor pool to be too sparse. Details of these imputation procedures are given in Appendix C.

---

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

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

In the 2002 NSDUH, because drug use was highly correlated with age, and to facilitate easier implementation of the imputation procedures, the model building and final assignment of imputed values for all drug use variables were performed separately within three distinct age groups: 12 to 17 year olds, 18 to 25 year olds, and persons 26 years of age or older.[37]

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

As with the CAI instruments used in the 1999 through 2001 surveys, the 2002 NSDUH had different drugs and drug use measures than are found in pre-1999 surveys. Exhibit 6.1 summarizes the drugs and drug use measures that were imputed and whether the imputations were univariate or multivariate. If no character is present in the box in Exhibit 6.1, then no information regarding that particular drug use measure was available for the given drug.

6.2 Hierarchy of Drugs and Drug Use Measures

The first step in the imputation process was to determine the order in which drugs and drug use measures were to be modeled, so that drugs and drug use measures earlier in the sequence could have been used, if applicable, as covariates for models fitted later in the sequence. Because the gate questions in the 2002 NSDUH were the basis for all subsequent drug data, it was necessary that the imputation of missing values for lifetime drug use for all drugs preceded imputations of all other drug use measures. These lifetime use indicators were temporary in the sense that they were manifested within the drug recency and frequency-of-use variables, but were 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 proceeded. As indicated in Exhibit 6.1, 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 30-day 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 had to be consistent (in a number of ways) with the other measures (see Section 6.5). Hence, age at first

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

use was imputed after the imputation for the other measures was completed.[38] The following sections describe the imputation procedures for each drug use measure.

Exhibit 6.1 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 values across all lifetime drug use variables.
× Multivariate neighborhood across recency of use, 12-month frequency of use where applicable, 30-day frequency of use where applicable, and the 30-day binge drink frequency variable (alcohol only); multivariate assignment of imputed values across measures.
×× Multivariate neighborhood across recency of use, 12-month frequency of use where applicable, and 30-day frequency of use where applicable; multivariate assignment of imputed values across these measures, and across certain drugs (see Section 6.4.5.1.2).
✓× 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, PCP, and Ecstasy.
[3] Includes methamphetamines.

6.3 Imputing Lifetime Drug Use Indicators

As with the 1999 through 2001 surveys, the 2002 NSDUH 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/her to the next module. Thus, the respondent was not

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

necessarily required to answer all questions in the questionnaire. The respondent could have skipped a module if he or she either indicated nonusage of the drug in the gate question or did not answer the gate question. Therefore, the gate question response was crucial to the range of responses available for subsequent questions in each module.

6.3.1 Hierarchy of Drugs

Since PMN was used for the lifetime usage imputations, a drug hierarchy was required, the use of which was motivated in general for PMN as described in Appendix C. Experience from past survey years indicated a substantial correlation between lifetime drug use indicators, as confirmed by an exploratory data analysis on the data from the 2002 NSDUH. Although models were built using respondents with complete data across all the drugs, predicted means had to be calculated for both item respondents and nonrespondents for lifetime use. When calculating the predicted means for the lifetime usage of a given drug for respondents who did not answer all the lifetime usage questions, a predictor value could have been missing. Hence, it was sometimes necessary to use imputed lifetime usage values. These imputed values needed to be provisional, since the final imputed lifetime usage indicators were not known until the final multivariate imputation, after the completion of the modeling.

Therefore, the first step in the imputation of lifetime indicators was to determine the order in which the drugs would be modeled, where drugs later in the sequence would have more predictors in their models. The order in which the lifetime indicators of use were imputed is shown in Exhibit 6.2.

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 have been considered a lifetime-use item respondent, he or she must have had complete data within each age group for all of the drug module gate questions: cigarettes, cigars, chewing tobacco, snuff, pipes, alcohol, marijuana, cocaine, crack, heroin, inhalants, LSD, PCP, Ecstasy, hallucinogens other than LSD, PCP, and Ecstasy, pain relievers, 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. (Because the modeling of the final weight adjustments was not completed at the time of the drug imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.)[39] 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

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

item response propensity model is a special case of the generalized exponential model (GEM),[40] which is described in greater detail in Appendix B.

Exhibit 6.2 Lifetime Indication of Use ("Gate") Questions (in Order of Imputation)[1]

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

[1] Follow-up questions were also considered in the lifetime imputation.
[2] Includes chewing tobacco and snuff.
[3] Includes LSD, PCP, and Ecstasy.
[4] Includes methamphetamines.

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

[40] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

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[41] was used to determine the parameter estimates. Because the interest was only in the estimation of the predictive mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. The predictors in each model included centered age,[42] centered age squared, centered 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, race, 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 E.

6.3.4 Computation of Predictive 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. In order to use lifetime usage of a given drug as a predictor for a drug later in the sequence, it was therefore necessary to utilize these temporary imputed values in cases where the original lifetime usage indicator was missing. If possible, provisional donors were chosen with predictive means within the delta of the recipient,[43] where the value of delta varied depending on the value of the predictive means, which in this case were predicted probabilities of lifetime use. In particular, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5, and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is given in Table 6.1. If no donors were available with predictive means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predictive mean was chosen.

---

[41] SAS®-callable SUDAAN® was used to fit the binomial and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in RTI (2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

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

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

Table 6.1 Values of Delta for Various Predicted Probabilities of Lifetime Use

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

6.3.5 Assignment of Provisional Imputed Values

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

6.3.6 Constraints on Univariate Predictive Mean Neighborhoods

In a general UPMN imputation, the neighborhood is restricted by two types of constraints: (a) logical constraints (which cannot be loosened) to make imputed values consistent with a nonrespondent's preexisting nonmissing values of other variables, and (b) likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible. 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 have also been considered a likeness constraint, which could have been 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 predictive 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 in the following sections.

6.3.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

Many respondents who indicated lifetime use of smokeless tobacco seemed to have been confused regarding the difference between chewing tobacco ("chew") and snuff, as was demonstrated by their responses to questions regarding specific brands. For example, many respondents who indicated use of chewing tobacco entered a snuff brand, such as Copenhagen™, when asked about the specific brand of chew they used. As a result, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted, rather than individual models for chew and snuff. The nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted probability of lifetime use. Missing values for chew and/or snuff were replaced with the values from a donor within this neighborhood. For individuals missing the lifetime usage indicator for either chew or snuff, but not both, only the missing value was replaced. However, for individuals missing both chew and snuff, both lifetime usage indicators were replaced by values from the same donor. No logical constraints were necessary in the assignment step. This was due to the fact that chew and snuff were assigned values independently, 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 2002 NSDUH questionnaire, separate models were fitted for the two substances. However, crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. This neighborhood was defined in terms of the deltas given in Table 6.1, which were 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 have been possible for a respondent to have been missing a value for cocaine, but not crack, because a crack user is, by definition, also a cocaine user. For this reason, no logical constraints were necessary.

6.3.7.3 Hallucinogens (LSD, PCP, Ecstasy, and Other Hallucinogens) 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 referenced in the subgate questions were of interest in their own right. For hallucinogens, there was interest in the usage of LSD, PCP, and Ecstasy; 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 with the exception of the ones of special interest. In the final assignment step, lifetime usage indicators were assigned for LSD, PCP, Ecstasy, and "other" hallucinogens, and for methamphetamines and "other" 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, PCP, and Ecstasy. It is important to note that if a respondent was a user of at least one of the other hallucinogens, he or she was considered a user of other hallucinogens, even if some of the other hallucinogens' subgate questions were unanswered. A missing value for other hallucinogens arose if at least one of the other hallucinogens' subgate questions was unanswered and all the other hallucinogens' subgate questions that were answered had a negative response. Using the neighborhood created from the hallucinogens' predicted probability of lifetime use, missing values for LSD and/or PCP and/or Ecstasy and/or other hallucinogens were replaced with the values from a donor within this neighborhood. For individuals missing a lifetime usage indicator for either LSD and/or PCP and/or Ecstasy and/or other hallucinogens, only the missing value(s) was (were) replaced. For individuals missing two or more of these lifetime usage indicators, the missing values were replaced by values from the same donor. As with smokeless tobacco, the subcategories for hallucinogens were assigned values separately, making logical constraints unnecessary. As a final step, a lifetime usage indicator for 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 those that dealt with methamphetamines, were all unanswered, or if these questions 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 2002 NSDUH were separated into item respondents and item nonrespondents for the lifetime drug variables. Subsequent sections summarize model building, computation of predictive means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predictive mean. In most cases, however, these univariate assignments were only provisional. As indicated in Exhibit 6.1, the final imputed values for these drug use measures were obtained by building neighborhoods upon a vector of predictive 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 have been a donor for a given item nonrespondent if he or she had complete data across all the lifetime drug use variables and was within the same age group.

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 missing only 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 predictive mean vector "close to" (i.e., within the delta distance of) the recipient's elements of the predictive mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predictive mean vector. The elements of the predictive mean vector corresponded to the predicted values of the recipient's missing lifetime use indicators. Initially, donors and recipients were required to have, if possible, the same values for all nonmissing lifetime use indicators. If this initial constraint did not produce a big enough donor pool, donors and recipients were 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 have been found within various likeness constraints is summarized in Appendix F. 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 predictive mean; and (4) remove the requirement that donors and recipients be from States with similar usage levels.

The full predictive mean vector contained elements for each lifetime drug use measure. However, only a portion of the full predictive mean vector was used; specifically, only those elements corresponding to the recipient's missing lifetime drug use were used. If the missing lifetime usage indicators corresponded to only one predictive mean, a UPMN imputation similar to the provisional UPMN was utilized. Otherwise, an MPMN imputation was employed. The Mahalanobis distance[44] was then calculated using only the portion of the predictive mean vector associated with the given missingness pattern. If no donors were available that had predictive means within a multivariate delta of the recipient's vector of predictive 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.

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

In the 2002 NSDUH, 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

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

frequency[45] 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 donors were found using the MPMN technique, even after loosening likeness constraints, UPMN values were used as final imputed values. (This was a safeguard that was never invoked for the 2002 survey.)

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

6.4.1 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 have arisen if too many covariates were included in the polytomous logistic models. (Lifetime usage indicators of other drugs were included instead of recency-of-use indicators.) However, for the sake of convenience, the recency of use imputations did follow the same hierarchy as described in Section 6.2.

6.4.1.2 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 reduced to lifetime users. Among these lifetime users, item respondents and nonrespondents for each drug were identified across recency of use and (where applicable) the 12-month, 30-day, and (for alcohol only) 30-day binge drinking frequency-of-use measures. If a valid response was provided for each drug use measure, the person was deemed an item respondent for the drug. Otherwise, he or she was an item nonrespondent.

Before modeling, the respondents' weights were adjusted so that they represented all lifetime users. (Weights were defined in the same way as with other drug use variables. See discussion about how the weights were defined in Section 6.3.2.) Because item respondents were defined at the drug level, these adjustments were made separately for each drug (and within the

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

three age groups). The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The covariates in the item response propensity model included age;[46] gender; race; first-order interactions gender and race; marital status; education; employment status,[47] census region; an MSA[48] 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.[49]

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, where possible, polytomous logistic regression.[50] The predictors included in the models were centered age;[51] centered age squared; centered age cubed; gender; race; first-order interactions of centered age, gender, and race; marital status; education; employment status;[52] 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 predictive mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. For a summary of the variables included in each drug model, see Appendix E.

For certain drugs, the proportion of users who were past year users was quite small when compared to the total number of lifetime users. The lopsided distributions[53] for these drugs caused convergence problems when fitting multinomial logistic models. This problem occurred with the following set of drugs that were either rare overall or were rare within one or more age groups: inhalants, hallucinogens, sedatives, stimulants, tranquilizers, and heroin. To alleviate this problem for these drugs for the 2002 survey, the single multinomial logistic model was

---

[46] The covariate age was divided into 5 categories to match the categories used in sample selection (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or over). For the 12-to-17 and 18-to-25 age groups, age was not included as a covariate in the item response propensity models.

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

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

[49] In a handful of cases (e.g., heroin, aged 26 or older), 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.

[50] SAS®-callable SUDAAN® was used to fit the polytomous logistic regression models. Details about the polytomous logistic regression model and additional references can be found in RTI (2001). SAS® software is a registered trademark of SAS Institute, Inc.; and SUDAAN® is a registered trademark of RTI International.

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

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

[53] A "lopsided distribution" in the context of recency of use is where, among the categories past month use, past year not past month use, and lifetime not past year use, only a tiny minority of respondents gave a response of "past month use."

replaced with two binary logistic models[54] that were fitted in a hierarchical manner. As with the multinomial logistic model, the first model was fitted among lifetime users, but the past month, and past year but not past month categories in the response variable were collapsed into a single level. In a similar manner to other recency-of-use models, respondents' weights were adjusted so that they represented all lifetime users. (Weights were defined in the same way as with other drug use variables. See the discussion about weights in Section 6.3.2.) Predictive means were obtained from the first model. Then, the second model was limited to past year users, where the response variable had two levels: past month and past year not past month users. For the second model, respondents' weights were adjusted so that they represented all past year users. (In order to do this, it was necessary to completely define the domain of past year users. Missing values were provisionally imputed to past year or not past year use by randomly allocating the response utilizing the predicted means from the first model.) From the two binary logistic models, the probability of past month use, and the probability of past year but not past month use were obtained and utilized in the final provisional UPMN, which is discussed in subsequent sections. Once the predicted means were determined from the two models, a single vector of predicted means conditional on lifetime usage, as with the multinomial logistic models, was determined in the following manner:

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

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

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

   P(past year, not past month use|past year use)**P*(past year use|lifetime use)

6.4.1.4 Computation of Predictive 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 predictive 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 appropriate logistic model(s). The predicted probabilities from the recency models were used to assign provisional values using the UPMN imputation method described in Appendix C. A vector of predicted probabilities for each respondent was created by the logistic regression model(s). Because only a single predictive mean was used to determine the neighborhood when determining provisional values, not all of the predicted probabilities from the model were used.[55] Also, because past month use was the most critical measure of recency of drug use, the neighborhoods were defined based on the probability of past month use. If possible,

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

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

provisional donors were chosen with predictive means within the delta of the recipient, where the value of delta varied depending on the value of the predictive means, which in this case were predicted probabilities of past month use.[56] 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 0 or 1. If no donors were available with predictive means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predictive mean was chosen.

6.4.1.5 Assignment of Provisional Imputed Values

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

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 have been within the same age group (12 to 17, 18 to 25, or 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. A small delta could have also been considered a likeness constraint, which could have been loosened by enlarging or removing delta. As previously stated, if no donors were found in the delta as defined in Section 6.4.1.4, the neighborhood was then abandoned, and the donor with the predictive mean closest to the recipient was chosen.[57] If possible, donors and recipients were required to be from States with the same level of usage of a given drug (the State rank, as defined in the introduction of this chapter), where the level of usage was defined in terms of the proportion of a given State's residents who had used a given drug in the past month. If insufficient donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned, and the donor with the closest predictive mean was chosen; (2) donors and recipients were no longer required to be from States with similar usage levels. 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 6.3. 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 have been impossible for a respondent to have last used the substance more than 3 years ago. Hence, under these circumstances, the donors were limited to having used within the past 3 years. Such a logical constraint would not have been useful for nontobacco products because the recency

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

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

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. Other logical constraints involving a very small number of respondents were not applied to the provisional imputations. The complete list of constraints used in the multivariate imputation of recency and frequency of use is given in Section 6.4.5.

6.4.1.7 Multivariate Assignments

Occasionally, more than one substance was associated with a single predictive mean, leading to a multivariate assignment of imputed values. However, for the provisional imputed values, a multivariate assignment was only necessary if the substances associated with a single predicted mean were of equal standing. This occurred with smokeless tobacco, which consists of chewing tobacco and snuff. No provisional imputed values were determined for substances that were a subset of the substance associated with the predicted mean. This occurred in modules that included subgate questions concerning substances that were of interest in their own right. These situations in the NSDUH were sometimes referred to as "parent/child" drugs, where the "parent" was the substance associated with the predicted mean, and the "child" was the subset substance. Examples of such situations included cocaine (parent) and crack (child), stimulants (parent) and methamphetamines (child); hallucinogens (parent); and LSD, PCP, and Ecstasy (children). The multivariate assignment of imputed values for chew and snuff is discussed below.

For reasons discussed in Section 6.3.7.1, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted rather than individual models for chew and for snuff. The nearest neighbor hot-deck neighborhood was then based on the predicted probability of past month use of smokeless tobacco. Missing recency-of-use values for chew and/or snuff were replaced with the (provisional) values from a donor within this neighborhood. At this stage in the process, lifetime use or nonuse of either chew or snuff was considered known (employing information from the lifetime usage imputation). For lifetime users of chew or snuff who were missing some or all of their recency-of-use information for either chew or snuff, but not both, only the missing specific recency-of-use values were replaced.[58] However, for individuals missing recency-of-use information for both chew and snuff (given that the respondent was known or was imputed to have been a chew user and a snuff user), values for both were obtained from the same donor. The provisional recency of use for smokeless tobacco was obtained by combining the recency-of-use information from snuff and chew.

---

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

Exhibit 6.3 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 (Nontobacco)
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	N/A
Lifetime, Not Past Year	1. Lifetime, not past 3 years 2. Past 3 years, not past year	N/A (for nontobacco, this was 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 have been past year users	Donors must have been past year users

6.4.2 12-Month Frequency of Use

6.4.2.1 Hierarchy of Drugs

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

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 item response propensity adjustment was then computed so that the respondents' weights accurately represented all past year users of the drug. (Weights were defined in the same way as with other drug use variables. See discussion about how the weights were defined in Section 6.3.2.) The item response propensity model is a special case of the GEM. The variables in the response propensity adjustment modeling included categorical 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.[59]

6.4.2.3 Model Building

As indicated in the previous section, only past year users of the drug of interest were used to build the 12-month frequency-of-use model. The (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 have been 100 / 365.25. Some respondents, however, started using the drug within the past year. If they responded to the month-at-first-use question, the difference between the month at first use and the date of the interview indicated the total time period during which they could have been using drugs.[60] 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 have been 100 / 101 instead of 100 / 365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

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

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:

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

where P_i is defined as the proportion of days in the past year in which respondent *i* used the drug. The standard logit transformation was not used because it was not defined for daily users. Using the adjusted weights, a linear univariate regression model using SUDAAN® software was then fitted for the log-transformed variable Y_i within each age group.

---

[59] If the recency of use for a particular drug was not yet defined, the lifetime indication of use was used instead. The recency of use of the drug being modeled (past month use versus past year but not past month use) was always defined.

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

Because the 12-month frequency models were limited to past year users, only two recency categories could have resulted: past month use and past year but not past month use.[61] Hence, recency of use for the drug being modeled was represented as a covariate in the 12-month frequency-of-use model by a single indicator variable representing these two categories. Imputation-revised recency of use for other drugs was used if available. If the missing values for a given drug's recency of use had not yet been imputed, a single covariate was used that indicated lifetime usage of that drug. To control for State variations in drug use, the State rank groups defined for the recency-of-use imputations were included as covariates in the 12-month frequency-of-use models.[62] Thus, the models included centered age;[63] centered age squared; centered age cubed; gender; race; State rank (based on past month prevalence of the drug); marital status; employment; education level;[64] census region; an MSA indicator; (where available) the imputation-revised 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 centered age, gender, and race.[65] Because interest focused only on the estimation of the predictive mean, and not on the parameter estimates exclusively or their standard errors, no model selection was attempted. Predicted 12-month frequencies of use were defined by back-transforming the resulting predicted values. For a complete summary of the 12-month frequency-of-use models, see Appendix E.

The predictive mean that resulted from the 12-month frequency-of-use model was a logit of the proportion of the year used. This logit was transformed back into a proportion for use as the variable from which the neighborhoods were created. This proportion could have been treated as a probability, which in turn could have been multiplied by the probability of past year use to make the predictive mean conditional on lifetime use of the drug in question. When calculating predictive means for some item nonrespondents, sometimes it was not known whether they were past year users. Hence, to make the predictive 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 Predictive Means and Univariate Predictive Mean Neighborhoods

Within a given drug, predictive 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 predictive mean conditional on lifetime use.

---

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

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

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

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

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

Using the UPMN methodology described in Appendix C, neighborhoods were defined based on these predictive means. If possible, provisional donors were chosen with predictive means within delta of the recipient, where the value of delta varied depending on the value of the predictive 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 0 or 1. As with recency of use, if no donors were available with predictive means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predictive mean was chosen.[66]

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. Also, assignments were not needed for "pills" because pills did not have a 30-day frequency-of-use question, making it unnecessary to obtain provisionally imputed 12-month frequencies. The final 12-month frequency-of-use imputations were multivariate across drug measures and are further described in Section 6.4.5.

6.4.2.6 Constraints on Univariate Predictive Mean Neighborhoods

An obvious logical constraint for 12-month frequency of use was that all donors were past year users. Other logical constraints involved the interview date, month of first use, birthday, recency of use, and 30-day frequency of use. See Section 6.4.5 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 predictive means of all potential donors were within 5 percent of the item nonrespondent's predictive mean, where the predictive mean was defined to be the proportion of the year (or maximum period within a year) during which a respondent used a drug. Finally, recipients and donors were required to have the same recency of use (past month versus past year but not past month),

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

whether that recency of use was reported or imputed.[67] 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 predictive 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 predictive mean. However, for the provisional imputed values, only the "parent" drug was of interest (for example, only the provisionally imputed cocaine 12-month frequency was needed and not the crack 12-month frequency). Therefore, multivariate assignments were not needed for the provisional UPMNs, but did occur in the final multivariate imputation of recency and frequency.

6.4.2.7 Multivariate Assignments

Although more than one substance was occasionally associated with a single predictive mean, the provisionally imputed 12-month frequencies were only required if they were needed for calculating predicted means using the coefficients from a subsequent model. A multivariate assignment was only necessary if the substances associated with a single predicted mean were of equal standing. This occurred with smokeless tobacco, which consists of chewing tobacco and snuff. However, no 12-month frequency was asked of smokeless tobacco users. Moreover, no provisionally imputed values were required for substances that were a subset of the substance associated with the predicted mean, which have been referred to as "parent/child" drugs (see Section 6.4.1.7). Hence, no multivariate assignments were required for the provisionally imputed 12-month frequency.

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 6.1).

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

The file was first reduced to the eligible population, which was past month users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the recency and 12-month frequency imputations. To have been an item respondent, the individual had to have provided valid responses to all applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of the drug. (Weights were defined in the same way as with other drug use variables. See the discussion in Section 6.3.2 about how the weights were defined.) The item

[67] Because all respondents in the 12-month frequency of use imputation were past year users by definition, item nonrespondents who were past month users required donors who were past month users, and item nonrespondents who were past year but not past month users required donors who matched that specific recency category.

response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. Predictors for the response propensity models included categorical 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 is apparent from the previous section, only past month users of the drug of interest were used to build the 30-day frequency-of-use model. The (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.[68] Using the adjusted weights, a linear univariate regression model was then fitted using SUDAAN® software for the log-transformed variable Y_i within each age group.

Because the 30-day frequency models were limited to past month users, only one provisional recency category was relevant for the drug of interest.[69] Hence, provisional recency of use for the drug of interest could not have been included in the 30-day frequency-of-use model. However, imputation-revised recency of use of other drugs could have been included. For drugs where the recency of use was not yet modeled, the lifetime indication of use served as a surrogate for the recency-of-use indicators. Covariates representing the State rank groups (defined by the level of past month use) were included to adjust for any State drug use differences. Other covariates included centered age;[70] centered age squared; centered age cubed; gender; race; marital status; employment; education level;[71] census region; an MSA indicator;

---

[68] 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 was defined for all respondents. (See Cox and Snell, 1989, for a discussion of the empirical logistic transformation.)

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

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

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

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 centered age, gender, and race. Because interest was only in the estimation of the predictive mean and not in the parameter estimates exclusively or their standard errors, no model selection was attempted. The predicted 30-day frequencies of use were defined by back-transforming the predicted values from the models. For a complete summary of the 30-day frequency-of-use models, see Appendix E.

The predictive mean that came out of the 30-day frequency-of-use model was a logit of the proportion of the month used. This logit was transformed back into a proportion for use as the variable from which the neighborhoods were created. This proportion was treated as a probability, which in turn was multiplied by the probability of past month use in order to have made the predictive means conditional on lifetime use of the drug in question. When calculating predictive means for some item nonrespondents, sometimes it was not known whether they were past month users or not. Hence, to make the predictive 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 this case, the response variable in a linear regression model was a logit of the proportion of the period (30 days) during which a respondent used the substance. The same pool of covariates was used in the logistic model and the regression model with the logit as the response variable. It should be noted that, unlike recency of use, the 30-day frequencies for snuff and chewing tobacco were not combined into a single value for smokeless tobacco. One could not have known if x days using snuff overlapped with the y days using chewing tobacco. Hence, separate models were fitted for snuff and chewing tobacco.

6.4.3.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods

Within a given drug, predictive means from the 30-day frequency-of-use models were computed for both item respondents and item nonrespondents using the parameters from the regression model. The 30-day frequency models were fitted after recency of use and 12-month frequency of use. The only drug for which provisional 30-day frequency values were required was alcohol because provisional 30-day frequencies were required to calculate 30-day binge drinking provisional values. Neighborhoods were created for each alcohol item nonrespondent using the UPMN technique described in Appendix C. The predictive 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)

For the 2002 NSDUH, an obvious logical constraint was that all donors had to have been 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 two likeness constraints were the three age groups and the State rank groups based on level of past month usage. As with the recency-of-use models, delta was set so that the predictive means of all potential donors were within 5 percent of the item nonrespondent's predictive mean, where the predictive 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 predictive mean was chosen; then (2) donors and recipients were no longer required to be from States with similar usage levels.

6.4.3.7 Multivariate Assignments

Although more than one substance was occasionally associated with a single predictive mean, the provisionally imputed 30-day frequencies were only required if they were needed for calculating predicted means using the coefficients from a subsequent model. Of the substances within the multivariate set of recency of use and frequencies of use, only alcohol contained a measure (30-day binge drinking frequency) that was lower in the sequence than 30-day frequency of use. Since alcohol is not a "parent/child" drug (see Section 6.4.1.7 for a definition of "parent/child" drug), no multivariate assignments were required for provisionally imputed 30-day frequency.

6.4.4 30-Day Binge Drinking Frequency

For alcohol, in addition to the 30-day frequency of use, an additional frequency variable was defined, which was 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[72] 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

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

binge drinking frequency measures; therefore, the weight adjustment used in the modeling of the 30-day binge drinking frequency was the same as was used for the 30-day frequency model.

The (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 binge drinking 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 binge drinking frequency for respondent i and N was the total number of days in the month that the respondent could have used the substance. This transformation was nearly equivalent to the standard logit transformation:

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

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

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

6.4.5 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, 6.4.3, and 6.4.4 summarize how the set of lifetime drug users in the sample of the 2002 NSDUH 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 predictive means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predictive mean. In most cases, however, these univariate assignments were only provisional. As is indicated in Exhibit 6.1, the final imputed values for these drug use measures were obtained by building neighborhoods upon a vector of predictive 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

---

[73] 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.)

question and was within the same age group. As with the provisional imputations, the donated value for the 12-month frequency of use variable was determined by taking the product of the donated proportion of the year that the donor had used the substance of interest and the recipient's maximum number of possible days that he or she could have used the substance.

6.4.5.1 Constraints on Multivariate Predictive Mean Neighborhoods

6.4.5.1.1 Logical Constraints

The logical constraints required in the provisional univariate imputations discussed in Sections 6.4.1, 6.4.2, and 6.4.3 were also required in the multivariate imputations. However, some constraints that potentially could have been applied in the provisional recency-of-use imputations were not applied because of the very small number of respondents affected, and are thus not listed in Exhibit 6.3. However, these constraints were applied in the multivariate imputations. In particular, the possible recencies of use were limited based on the respondent's current age, the time between the interview date and the birthday, the time between the interview date and the month of first use, and any nonmissing frequency-of-use information. In general, the application of these constraints depended on what information was missing in the recency-of-use and frequency-of-use variables. The values missing for a given respondent define the "pattern of missingness." For example, one pattern of missingness for marijuana could be as follows: past year user of marijuana (recency partially missing), 12-month frequency not missing, and 30-day frequency missing. In this example, the logical constraints have to make the imputed 30-day frequency consistent with the preexisting 12-month frequency. In the case where the 12-month frequency of use variable was missing, an additional logical constraint involved the product of the donated proportion and the recipient's maximum possible number of days used in a year (called the "donated 12-month frequency product"). Since this product involved both the donor and the recipient, it had to be consistent with the 30-day frequency of use, regardless of whether the 30-day frequency was a preexisting nonmissing value or a donated value. 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 G.

6.4.5.1.2 Likeness Constraints

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

6.4.5.1.3 More than One Substance for a Single Predicted Mean

Occasionally, more than one substance was associated with a single predictive mean, whether it was for recency of use or the frequency-of-use variables. This could have been two substances of equal standing considered together when modeling (snuff and chewing tobacco) or drugs with a parent/child relationship (see Section 6.4.1.7 for a definition of parent/child relationship). The assignment of imputed values for these substances was unique for each situation. Hence, the imputations for each of these substances are discussed as follows.

Smokeless Tobacco

As noted in Sections 6.3.7.1 and 6.4.1.7, 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 predicted probability of past month use of smokeless tobacco. The assignment of recency-of-use values for smokeless tobacco followed the same logical constraints in the multivariate imputation as was the case for the univariate imputations discussed in Section 6.4.1.7.

Unlike recency of use, however, separate models for snuff and chew were built for 30-day frequency of use. The predictive 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 predictive means used to form the neighborhoods were conditioned on lifetime usage rather than past month usage. Because the 30-day frequency models gave predictive means conditioned on past month use, it was necessary to determine the probability of past month use given lifetime use, which could have been obtained from the recency models. Because the 30-day frequencies for snuff and chew could not have been 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.) The covariates used in the models are the given in Appendix E.

Cocaine and Crack

Even though cocaine and crack were in distinct modules in the 2002 NSDUH questionnaire, single models were fitted for recency of use and the frequency-of-use variables using the information from the cocaine module. Crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. As with smokeless tobacco, use or nonuse of crack was considered known (using information from the lifetime imputations). Hence, as a logical constraint, users of crack with incomplete recency (or frequency) information required donors who were also crack users. Moreover, if the cocaine recency was not missing, the donated crack recency could not have been more recent than the preexisting cocaine recency. Similarly, if the crack recency was not missing but the cocaine recency was missing, the donated cocaine recency could not have been less recent that the preexisting crack recency.

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

Hallucinogens (LSD, PCP, Ecstasy, and Other Hallucinogens) 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, PCP, and Ecstasy; 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, Ecstasy, 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 not necessary. 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, Ecstasy, and methamphetamines was considered known (including values that were imputed in 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 Ecstasy and/or hallucinogens, as a whole, were replaced with the specific recency categories from a single donor. LSD, PCP, and Ecstasy users with incomplete recency information were constrained to have donors who were LSD, PCP, and Ecstasy users, respectively. Moreover, donors were constrained so that a preexisting LSD, PCP, or Ecstasy recency could not have been more recent than a donated hallucinogens recency; conversely, a preexisting hallucinogens recency-of-use value could not have been less recent than donated LSD, PCP, or Ecstasy recency of use. For individuals missing recency information for either LSD and/or PCP and/or Ecstasy 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.

No 12-month frequency-of-use variables were available for LSD, PCP, or Ecstasy; however, the "donated 12-month frequency product" for all hallucinogens had to be consistent with the 30-day frequency-of-use value for all hallucinogens, whether it was imputed or was a preexisting nonmissing value.

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-of-use value could not have been more recent than a donated stimulant recency-of-use value, and conversely, a preexisting stimulant recency-of-use value could not have been less recent than donated methamphetamine recency of use. For individuals missing recency information for methamphetamines and/or stimulants, as a whole, only the missing categories were replaced. For individuals missing recency information on both of these substances, the missing categories were replaced by values from the same donor.

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

6.4.5.2 Final Multivariate Assignment

The full predictive mean vector contained several elements for recency of use (different probabilities associated with each of the recency categories), as well as elements for the frequency-of-use variables. Each element in the full vector of predictive means was adjusted so that all elements were conditioned on the same usage status whenever possible. The resulting elements in the predictive mean vector that could have potentially resulted are given in Exhibit 6.4. It is important to note that not all drugs contained all the elements given. Exhibit 6.5 shows the full predictive mean vector for each drug. The portion of the full predictive mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If partial information was available regarding recency of use, that information was used to adjust the recency-of-use probabilities. The portions of the full predictive mean vector used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given in Appendix G. The Mahalanobis distance was then calculated using only the portion of the predictive mean vector that was associated with the given missingness pattern, with elements appropriately adjusted. If no donors were available that had predictive means within a multivariate delta of the recipient's vector of predictive 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.

Exhibit 6.4 Elements of Full Predictive Mean Vector

Drug Use Measure and Category of Interest	Predictive 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) was not needed in the predictive mean vector because the multinomial probabilities added to 1, and this probability was determined by the other probabilities.

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

Exhibit 6.5 Full Predictive Mean Vector for Sample Drugs

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

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

6.5 Age at First Use and Related Variables

Unlike the recency and 12-month frequency-of-use variables, age at first drug use was not statistically imputed in the surveys prior to the 1999 survey; instead, missing values were excluded from subsequent analyses. However, as with the 30-day frequency, missing age-at-first-use values have been imputed since the 1999 survey. Also, recent drug initiates (i.e., those whose current age was equal to or 1 year greater than the reported age at first use) were asked the year and month of their first use. To have this information for all users, both missing year and missing month of first use for less recent initiates (and recent initiates who did not report year and month of first use) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised age at first use, and imputation-revised recency and frequency variables. To have complete date of first use information, day of first use was randomly assigned for all users. The combined data gave the respondent's age at first use along with the date of first use. It is important to note that in addition to age at first use for cigarettes, those respondents classified as lifetime daily cigarette users 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 utilizing SUDAAN® software was fitted using a logit transformation of the respondent's age at first drug use as the response variable. UPMNs were formed using the predictive mean from the regression model. Each item nonrespondent's neighborhood was restricted by logical constraints (which could not have loosened) and likeness constraints (which could have been 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 and recency/frequency-of-use usage hierarchy given in Exhibit 6.2.

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

6.5.1.3 Sequential Model Building

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

$$Y_i = \ln[p_i/(1-p_i)], \text{ where } p_i = \frac{Age\ of\ First\ Use_i + Uniform(0,1)\ Number}{(Interview\ Date - Date\ of\ Birth + 1)\ (365.25)}$$

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[74] were centered age;[75] centered age squared; centered age cubed; State rank (based on the recency variable, see Section 6.4.1 for details); gender; race/ethnicity; first-order interactions of centered age, centered age squared, gender, and race/ethnicity; marital status; education level; employment status;[76] 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 (where available, otherwise lifetime indicators were used); 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:

$\mathbf{new12_i}$	=	0	if respondent did **not** use the i^{th} drug in the past 12 months
	=	12-month frequency	if respondent used the i^{th} drug in the past 12 months
$\mathbf{new30_i}$	=	0	if respondent did **not** use the i^{th} drug in the past month
	=	30-day frequency	if respondent used the i^{th} drug in the past month
$\mathbf{AFU_i}$	=	0	if respondent is **not** a lifetime drug user of the i^{th} drug
	=	age at first use	if respondent is a lifetime drug user of the i^{th} drug

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 E.

---

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

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

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

6.5.1.4 Computation of Predictive 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 "predictive mean" was the predicted age at first use. Again, this procedure defined a "neighborhood" of respondents by requiring that the respondents' predicted age-at-first-use values 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 available with predictive means within 5 percent of the recipient's predictive mean, the neighborhood was abandoned, and the respondent with the closest predicted age at first use was chosen as the donor.

6.5.1.5 Assignment of Imputed Values

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

6.5.1.6 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 years, 18 to 25 years, or 26 years or older). Models were built separately within these three groups, so this likeness constraint was never loosened. In fact, recipients and donors were required to be of the same age, if possible. If a donor of the same age was not found, the constraint eventually reduced to a logical constraint, where the imputed age at first use was less than the recipient's age. A small delta could have also been considered a likeness constraint, which could have been loosened by enlarging or removing delta. Initially, the relative distance for determining age at first use imputation neighborhoods (delta) was set so that any potential donor's predicted age-at-first-use was within 5 percent of the recipient's predicted age at first use, 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 occurred as follows: (1) remove the State rank group; (2) abandon the neighborhood, and choose the donor with the closest predictive 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 have been 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;[77] and (5) loosen the "same-age" restriction even further, so that the donor's age at first use could have been less than or equal to the recipient's age. A summary of the above constraints and the number of respondents who fitted into each one are listed for each drug in Appendix F.

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 is 90. If the respondent had a 12-month frequency of 100 (either reported or imputed), his or her age at first use could not have been his or her current age. Second, if the respondent's recency of use indicated that he or she did not use in the past month, but the number of days since his or her last birthday was fewer than 30, the recipient's age at first use could not have been equal to his or her current age. And third, if the respondent was not a past month user, but the difference between his or her 12-month frequency and the days since his or her last birthday was fewer than 30, the recipient's age at first use could not have been equal to his or her current age. Consider again the example where the recipient respondent's birthday was on March 1st, 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 recency of use indicated past year but not past month use, 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 have equaled his or her current age. Some additional logical constraints were that the donors could not have been past year users if the recipient was not a past year user, and, for tobacco, donors could not have been 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. Starting with the 2002 survey, respondents with age of first use values of 1 or 2 were no longer eligible to be donors. In addition, if the editing indicators for cigarettes, hallucinogens, stimulants, or cocaine denoted that logical editing occurred, these respondents were also not eligible to have been donors. Finally, cigarettes had yet another logical constraint: if the recipient was a daily cigarette user and his or her age at first daily use was not missing, the donors were prevented from having an age at first use later than the preexisting age at first daily use.

6.5.1.7 Multivariate Assignments

For smokeless tobacco (chewing tobacco and snuff), cocaine (crack), hallucinogens (LSD, PCP, and Ecstasy), and stimulants (methamphetamines), more than one age-at-first-use variable was associated with a single predicted age at first use. This led to a multivariate assignment of the imputed values. Drugs where multivariate assignments were necessary are discussed in the following sections.

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

6.5.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff)

For reasons discussed in Section 6.3.7.1, one model for smokeless tobacco was fitted rather than individual models for chewing tobacco and for snuff. The nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted age at first use. Missing age-at-first-use values for chewing tobacco and/or snuff were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and if both chewing tobacco and snuff were missing, imputed values came from the same donor. As for the constraints on the neighborhoods, all the constraints listed in the previous section were applied to both snuff and chewing tobacco separately. 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 have been 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 have been adjusted to reflect the situation. Age at first use for smokeless tobacco was obtained by taking the minimum age at first use from snuff and chewing tobacco.

6.5.1.7.2 Cocaine and Crack

Even though cocaine and crack were in distinct modules in the 2002 NSDUH questionnaire, an age-at-first-use model was only fitted 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 have been 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 have been adjusted to reflect the situation.

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

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

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

Because LSD, PCP, and Ecstasy are all types of hallucinogens, additional logical constraints were required so that donated values were consistent with preexisting nonmissing values. For example, if the age at first use for LSD and PCP were missing but overall age at first use hallucinogens and Ecstasy were not, the donated LSD and PCP age at first use could not have been earlier than the preexisting hallucinogens' age at first use (however, the LSD and PCP age at first use could have been earlier than the Ecstasy age at first use). Another example is if the age at first use for hallucinogens was missing and the LSD age at first use was not (and the respondent was a nonuser of PCP and Ecstasy), then the donated hallucinogens' age at first use could not have been later than the preexisting LSD age at first use. In addition, if the LSD, PCP, or Ecstasy age at first use was missing, but the respondent was a user, the donor had to have matched the respondent's lifetime usage pattern. Finally, if the respondent used LSD and/or PCP and/or Ecstasy, but used no "other" type of hallucinogen, then their overall hallucinogen age of first use was imputed (or assigned) to be consistent with the minimum age of first use of LSD and/or PCP and/or Ecstasy.

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

6.5.1.7.4 Stimulants (Methamphetamines and Other Stimulants)

As stated in Section 6.3.7.3, the stimulants module included a subgate question referring to methamphetamines, which was of interest in its own right. One model was fitted 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 methamphetamines are a type of stimulant, additional logical constraints were required so that donated values were consistent with preexisting nonmissing values. Specifically, if the age at first use for methamphetamines was missing but overall age at first use for stimulants was not, the donated methamphetamines' age at first use could not have been 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 have been later than preexisting methamphetamines' age at first use. In addition, if the methamphetamines' age at first use was missing but the respondent was a methamphetamines user, the donor had to have been a methamphetamines user. Finally, if the respondent used methamphetamines, but used no "other" type of stimulant, then their overall stimulant age of first use was imputed (or assigned) to be the same value as their methamphetamine age of first use. 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 have been 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 instrument: occurs when AFU < current age - 1).

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

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

4. For recent initiates: Missing day of first use only (asked in 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 data prior to the 1999 survey, 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*) * (a random number generated from a Uniform(0,1) distribution)],

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 or 365)][78] *if recency* = 1

minimum [(Interview date - 29 - 12-month frequency), (Earliest date + 364 or 365)] *if recency* = 2

minimum [(Interview date - 1 day - 1 year), (Earliest date + 364 or 365)] *if recency* = 3

minimum [(Interview date - 1 day - 3 years), (Earliest date + 364 or 365)] *if recency* = 4

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

[78] The number added to "earliest date" was set to 364 if the interview date was a nonleap year and it was set to 365 if the interview date was a leap year.

MFU and DFU were derived in the same manner as the date of first use in Missingness Pattern 1 with the following changes:

- If the *Earliest possible date* < YFU, then *Earliest date* = YFU (using January 1st as the earliest month/day).
- If the *Latest possible date* > YFU, then *Latest date* = YFU (using December 31st as the latest month/day).

6.5.1.8.3 Missingness Pattern 3

Similar to Missingness 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 Missingness 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 Missingness 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 or 365)] *if recency* = 1

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

minimum [(Interview date - 1 day - 3 years), (Earliest date + 364 or 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 parent/child relationship drugs (i.e., cocaine and crack; stimulants and methamphetamines; hallucinogens and LSD, PCP, and Ecstasy) 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 and hallucinogens/LSD/PCP/Ecstasy). Moreover for stimulants/methamphetamines, if the respondent used methamphetamines but no "other" stimulant, then the stimulant date of first use was assigned to be equal to the methamphetamines date of first use (if possible).[79]

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 daily cigarette use, where a daily user was defined as someone who reported having at some time smoked cigarettes every day for a period of at least 30 days. Imputation procedures for age at first cigarette daily use were similar to age at first use, with 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 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 have been imputed.

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

6.5.2.1 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 years, 18 to 25 years, and 26 years or older), and all subsequent procedures were performed separately

[79] The same logic applies to hallucinogens/LSD/PCP/Ecstasy.

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[80] 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: categorical age, race, gender, census region, an MSA indicator, and imputed recency of use for cigarettes and the lifetime indicators of cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

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 in SUDAAN®. The predictive mean from this model was the predicted probability of ever smoking cigarettes daily. Variables included in the initial regression equation were centered age; centered age squared; centered age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of centered age, centered age squared, gender, and race/ethnicity; marital status; education level; employment status;[81] census region; an MSA indicator; imputed recency of use for cigarettes and the lifetime indicators for 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 E.

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

From the final model, a predictive 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 "predictive 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 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.

---

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

[81] 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 have been within the same age group (12 to 17 years, 18 to 25 years, or 26 years or older). Models were built separately within these three groups, so this likeness constraint was never loosened. The likeness constraints were nearly identical to those of age at first use (see Section 6.5.1.6). The only difference was in the definition of the predictive mean, the determination of which was described in Section 6.5.2.2. A summary of the likeness constraints and the number of respondents who fitted into each one are listed for each drug in Appendix F.

6.5.2.6 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 incorporated 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 they smoked everyday for at least 30 days.[82] The file was broken down into three age categories (12 to 17 years, 18 to 25 years, and 26 years or older), and all subsequent procedures were performed separately within these age groups. 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 and the lifetime indicators for 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 where the dependent variable underwent 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 centered age; centered age squared; centered age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of centered age, gender, and race/ethnicity; marital status; education level; employment status;[83] census region; MSA; imputed recency of use for cigarettes and the lifetime indicators for 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 E.

---

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

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

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

From the final model, a predictive 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 predictive mean. The imputation-revised age-at-first-daily-use assignment was conducted using UPMN imputation. The procedure defined a "neighborhood" of respondents by requiring that the respondent's predicted 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 years, 18 to 25 years, or 26 years or older). Models were built separately within these three groups, so this likeness constraint was never loosened. The likeness constraints were nearly identical to those of age at first use (see Section 6.5.1.6). There were only two differences. First, the predictive mean was defined differently, as described in Section 6.5.2.6. Second, an additional step was employed if no donor was found after loosening all of the likeness constraints. In particular, if the age at first use and age at first daily use were both initially missing, the imputed age at first use was set back to missing, and reimputed simultaneously with the age at first daily use, so that both were mutually consistent.[84] A summary of the above constraints and the number of respondents who fitted into each one are listed for each drug in Appendix F.

All the logical constraints applied to age at first cigarettes' use were also applied to age at first daily cigarette use; in other words, simply replace the words "age at first use" with "age at first daily use" in Section 6.5.1.6. 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 age at first daily cigarette use was created, all daily cigarette users had a valid age of first daily cigarette use. From this age, a year/month/day of first daily use was assigned. Starting in the 2002 NSDUH, respondents were asked their daily month and year of first daily use (before the 2002 survey, respondents were only asked their age of first). Due to this change in the questionnaire, new dates of first daily use variables were created to reflect this additional information provided by the respondent (IRCD2YFU, IICD2YFU,

[84] The situation described here did not occur in the 2002 survey.

IRCD2MFU, IICD2MFU, IRCD2DFU, and IICD2DFU). The date assignment procedure was identical to the missingness patterns described in Section 6.5.1.8. An additional constraint was required that ensured that the date of first daily use was on or after the date of first cigarette use.

In previous survey years, the date of first daily cigarette use variables were created under the assumption that data on the respondent's year and/or month of first daily use of cigarettes was not available. These variables (IRCDUYFU, IICDUYFU, IRCDUMFU, IICDUMFU, IRCDUDFU, and IICDUDFU) were created in the 2002 NSDUH to allow for consistency between survey years. The assignment procedure for these variables was similar to Missingness 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*) * (a random number generated from a Uniform(0,1) distribution)]

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

7. Nicotine Dependence Imputations

7.1 Introduction

The questions concerned with nicotine dependence in the 2002 National Survey on Drug Use and Health (NSDUH)[85] were the same as those asked in the 2001 survey, when a new way of measuring dependence on nicotine through cigarettes, clove cigarettes, or bidis[86] was introduced. In particular, this method involved the calculation of a continuous scale of nicotine dependence, called the Nicotine Dependence Syndrome Scale (NDSS) (Schiffman et al., 1995; Schiffman et al., 2003). This scale was calculated from 17 NSDUH questionnaire items (see Exhibit 7.1), which were asked of respondents who used cigarettes in the past 30 days. For a response to be considered valid, an answer of either "1=Not at all true," "2=Somewhat true," "3=Moderately true," "4=Very true," or "5=Extremely true" had to be given to each of the 17 questions. The scale was the mean value (appropriately adjusted where necessary) of the responses to the 17 questions, provided all 17 responses were nonmissing.

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

The imputations described in this chapter are almost unique in this report due to the fact that they were not performed using the predictive mean neighborhood (PMN) technique as described in Appendix C. Another exception to the PMN method was the imputation of missing values for the immigrant status variables, which used a weighted hot-deck procedure as described in Chapter 5. It should also be noted that the NDSS mean value was calculated from edited versions of the 17 nicotine-dependence questionnaire variables. The majority of the editing procedures for these variables are described elsewhere (Kroutil, 2003a, 2003b, 2003c).

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

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

7.2 Edited Nicotine Dependence Variables

Exhibit 7.1 shows the correspondence between the 17 questionnaire items used in the NDSS and the corresponding edited variables. Among eligible respondents (those who had used cigarettes, clove cigarettes, or bidis in the past 30 days), the valid responses for the edited variables, as with the raw variables, were given as "1=Not at all true," "2=Somewhat true," "3=Moderately true," "4=Very true," or "5=Extremely true" had to be given. For most nicotine dependence variables, "dependence" was marked by the "Extremely true" response. However, for question variables DRCGE04, DRCGE12, DRCGE13, and DRCGE14, "dependence" was marked by "Not at all true."

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

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

7.3 Imputed Nicotine Dependence Variables

7.3.1 Setup for Model Building

In general, imputation models for variable types other than nicotine dependence in the 2002 NSDUH were modeled sequentially, so that variables that were modeled early in the sequence could have been used as covariates in models for variables later in the sequence. This was done to avoid fitting separate models for each missingness pattern. In the case of nicotine dependence, however, no imputation was performed if more than one NDSS variable was missing. As a result, for each respondent where imputation could have been performed, all 16 nonmissing NDSS variables could have been used as covariates in the model for the 17^{th} missing variable. Therefore, no sequential modeling was necessary. Item respondents therefore had to have complete data for all 17 of the NDSS questions used in the models, and logically they had to have used cigarettes, clove cigarettes, or bidis in the past 30 days. Item nonrespondents were those who used cigarettes, clove cigarettes, or bidis in the past 30 days and answered only 16 of the 17 NDSS questions with valid nonmissing responses. Respondents, who had used cigarettes, clove cigarettes, or bidis in the past 30 days and were therefore eligible to answer the NDSS questions, but only answered 15 or fewer of those questions, were left out of the modeling process. The missing values in the NDSS variables for these respondents remained missing in the imputation-revised variables. No response propensity adjustments were performed for the item respondent weights used in any of the models. However, the ratio-adjusted design-based weights were used in the imputation models. The variables included in the models are discussed in the next section.

7.3.2 Model Building

In the 2002 NSDUH, one model was created for each NDSS variable. The response variable for each model was the edited variable that corresponded to the question text given in Exhibit 7.1. The covariates in each model were the remaining NDSS variables. For example, if CIGIRTBL was the response variable, then the covariates would be the remaining 16 NDSS variables: CIGCRAVE, CIGCRAGP, CIGINCTL, CIGAVOID, CIGPLANE, CIGRNOUT, CIGREGDY, CIGREGWK, CIGREGNM, CIGNMCHG, CIGSVLHR, CIGINFLU, CIGNOINF, CIGINCRS, CIGSATIS, and CIGLOTMR.

7.3.3 Computation of Predictive Means

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

7.3.4 Assignment of Imputed Values

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

7.4 Summary Information for Nicotine Dependence Variables

Imputations were necessary for the nicotine dependence variables to create an NDSS score for as many eligible people as possible. The imputation method was devised to be simple and easy to implement, given the complexities of handling this type of missing data. To avoid complicated models, imputations were limited to cases where the respondent answered 16 of the 17 questions. If an eligible respondent answered fewer than 16 questions, no imputations were performed. In fact, in some cases, the eligibility to answer the NDSS questions was not clear. Specifically, this was possible in the case where a respondent was not a past month user of cigarettes and did not answer at least one of the bidis and clove cigarettes past-month-use questions. In these unclear cases, the respondent could have been eligible if the unknown missing response(s) for at least one of the questions had implied usage of bidis or clove cigarettes in the past month. It was also possible that the respondent was only eligible to answer the question because he or she was imputed to be a past month cigarette user, and the bidis/clove cigarette questions were not answered affirmatively.[87] Exhibit 7.2 summarizes the eligibility of respondents to answer the nicotine dependence questions and reasons why the respondent was eligible or not eligible. Furthermore, among respondents who were eligible, this exhibit gives details about the amount of nicotine dependence data that was missing. Also, this exhibit provides information on whether the respondent was imputed to be a past-month cigarette user; consequently, the respondent would have been eligible to have nicotine dependence data, but would have had missing data for all the nicotine dependence variables.

[87]It was possible for an imputed past month user with missing cigarette dependence data, to have had raw cigarette dependence data available. These raw dependence data would have been set to bad data if the respondent was initially a past month user, but were edited to a broader recency category. The nicotine dependence data were set to bad data because they were time-dependent. The final imputation-revised variable did not incorporate these raw data.

Exhibit 7.2 Summary of Response Patterns for NDSS Variables

Number of Missing NDSS Variables	Past Month Smoker	Past Month User Bidis or Cloves	Frequency
NOT ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS: 48,498			
17	no (not imputed)	no	47,899
17	no (imputed)	no	599
ELIGIBILITY TO ANSWER NICOTINE DEPENDENCE VARIABLES UNKNOWN: 432			
17	no (not imputed)	not known	426
17	no (imputed)	not known	6
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE QUESTIONS, MISSING VALUES IN DEPENDENCE VARIABLES NOT IMPUTED: 169			
17	yes (not imputed)	no or not known	12
17	yes (imputed)	no or not known	16
17	no (not imputed)	yes	0
17	yes (not imputed)	yes	0
2-16	yes (not imputed)	no or not known	127
2-16	no (imputed or not imputed)	yes	10
2-16	yes (imputed or not imputed)	yes	4
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE VARIABLES, MISSING VALUES IN DEPENDENCE VARIABLES IMPUTED: 224			
1	yes (not imputed)	no or not known	203
1	no (imputed or not imputed)	yes	4
1	yes (not imputed)	yes	17
KNOWN ELIGIBLE TO ANSWER NICOTINE DEPENDENCE VARIABLES, NO MISSING VALUES IN DEPENDENCE VARIABLES: 18,803			
0	yes (not imputed)	no or not known	17,476
0	no (imputed or not imputed)	yes	254
0	yes (imputed or not imputed)	yes	1073

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 for the 2002 National Survey on Drug Use and Health (NSDUH).[88] As with 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 are described elsewhere (see Kroutil 2003a), 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 household roster portion of the questionnaire (QD54) was interviewer administered. This question asked the respondent for information regarding the number of people living in his or her household, where allowable entries ranged from 1 to 25. If either the interviewer indicated that the respondent lived alone or the question was unanswered, the household composition (roster) section was skipped. However, if the interviewer indicated a household size greater than 1, the interviewer was then prompted to ask the respondent questions about the age, gender, and relationship to the respondent of every member of the household, starting with the household's oldest member, and including the respondent. If a pair of respondents was selected in a household, the interviewer indicated which member of a respondent's household roster corresponded to the other selected pair member. The roster entry for the respondent was referred to as the "self" entry. In effect, the respondent 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 8.1. 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 8.2. Also given in Exhibit 8.2 are details corresponding to certain relationship codes.

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

8.2.2 Household Roster Consistency Checks

To reduce the amount of editing required during the data processing stage, consistency checks were included in the Blaise program code.[89] Two types of consistency checks were employed in the household roster section of the questionnaire. These checks (1) compared a roster entry corresponding to the respondent with previously entered questionnaire information or (2) checked a roster entry against other roster entries or the respondent's roster age for internal consistency.

8.2.2.1 Comparisons with Previously Entered Questionnaire Information

In the 2001 NHSDA, a consistency check was added that was triggered if the interviewer reported a gender for the respondent in the household roster that did not match the gender entered in the beginning of the interview (QD01, the first question in the questionnaire). The interviewer was required to change either the roster entry or the gender that had been entered at the beginning of the interview. In the 2002 NSDUH, a new consistency check involving the respondent's age was added. Not only was it necessary for the respondent's gender in the roster to match the questionnaire gender, but also for the respondent's age in the roster to match the age that had been entered in the non-roster part of the questionnaire (the Blaise variable CURNTAGE). For the age check, however, the interviewer could have either changed the respondent's age entered in the roster, or overridden the consistency check and provided an explanation as to why the roster age did not match CURNTAGE. Both of these consistency checks necessarily involved the respondent's own entry in the roster (the "self" entry). If the consistency check for age was overridden, the value for age corresponding to the self may not have matched the questionnaire-edited age. Strategies employed for this situation are discussed in Section 8.2.4. Explanations given by the interviewer for overriding this particular consistency check were not reviewed.

8.2.2.2 Internal Consistency Checks

New consistency checks that were added for the 2002 survey also checked for internal consistency in the roster. These checks were triggered if:

1. the interviewer reported that the respondent had more than one spouse of the same gender. This check was not applied for multiple spouses of different genders, nor was it applied for live-in partners, or any combination of live-in partners and a single spouse.
2. the interviewer reported that a household member was a parent or grandparent of the respondent and the respondent was older than the household member.
3. the interviewer reported that a household member was a child or grandchild of the respondent and the respondent was younger than the household member.

In most cases, if the consistency check was triggered, the interviewer changed either an age or a relationship code in the roster to a more appropriate value. However, as with the check of the respondent's age, if the consistency check was overridden, the interviewer's explanations

---

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

had to be checked individually. If a given explanation was considered legitimate (e.g., a father marries a woman, listed as [step]mother, who is younger than the respondent), the unusual response was allowed to stand. Otherwise, the relationship code was set to bad data. Explanations given by the interviewers for the overrides, and evaluations of their legitimacy, are given in Appendix I.

8.2.3 Preliminary Roster Edits

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

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

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

[1] This only applied to respondents who were part of a pair. The other member selected could not have been the self because respondents were not interviewed twice. The other member selected was the roster member who had a value of "1" for this variable.

Exhibit 8.2 Household Composition (Roster) Relationship Codes

Relationship Code #	Relationship to Respondent	Details About Relationship
1	Self	
2	Parent	Biological, Step, Adoptive, or Foster
3	Child	Biological, Step, Adoptive, or Foster
4	Sibling	Full, Half, Step, Adoptive, or Foster
5	Spouse	
6	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	
15	Marked as (Live-in) Partner but Not Possible	

8.2.4 Roster Edits Involving the Self

The Blaise program code required the interviewer to identify exactly one "self" in the household roster. When the interviewer identified the self, he or she also applied an age and gender, which should have, in theory, matched CURNTAGE and QD01, respectively. Because the check involving gender was not allowed to be overridden, the gender for self in the roster always matched QD01, which was equivalent to IRSEX (see Chapter 4). Given the consistency check comparing the respondent's roster age against CURNTAGE, the age of self in the roster should at least have been close to the questionnaire-edited age, AGE (see Chapter 4 for a description of the methodology used to create AGE). Nevertheless, since an interviewer could have overridden the consistency check, it was possible, in rare instances, to have had problems matching AGE with the age of self in the roster. The interviewer overrode the consistency check for age of self for one of two reasons: (1) the self was misidentified, and another roster member was the true self, but the interviewer insisted on not changing the entries, or (2) the interviewer correctly identified the self, but insisted that the correct age for the respondent was different than CURNTAGE. In the case of a misidentified self, a second roster member in the household was selected as the self whose gender matched IRSEX, and whose age was within one year of AGE. For the latter case, no such roster member was found. In these instances, the roster member identified as self was replaced with a self that had an age and gender that matched IRSEX and AGE.

If the consistency check was overridden, a misidentified self was diagnosed if (1) the roster age of self differed from AGE by more than 1 year, and (2) another roster member of the same gender as QD01 (and IRSEX) had a roster age within one year of AGE.[90] Assuming a misidentified self, the interviewer used the roster member identified as the self, rather than the respondent, as the point of reference. Using the example given in Exhibit 8.1, if the respondent's son was used as the reference point, the relationship for the respondent became "mother" instead of "self" and the "husband" became "father." Under these circumstances, the code for self was set to missing, and the respondent's roster entries did not include a self. The remainder of relationship codes in the roster was also set to missing. In some cases, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point.

8.2.4.1 Original Self Misidentified: Identifying the Real Self

If the self was misidentified in the roster, an attempt was made to identify a self among the roster members corresponding to the respondent 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. If more than one roster member met the above criteria, the roster members who met the criteria, but were not assigned the self code, were given a bad data code.

[90] A 1-year difference was allowed since the respondent's age might have changed during the interview. In this instance, the values of AGE and CURNTAGE may have differed by 1 year.

8.2.4.2 Salvaging Relationship Codes with a Misidentified Self

As stated earlier, if the self was misidentified, all other relationship codes were set to missing because the reference person was someone other than the respondent. In some cases, however, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point. Relationship codes were salvaged under the following circumstances:

1. If the reference person was the respondent's sibling, the roster member listed as "self" was actually a sibling, and all other relationship codes were salvaged. (Presumably, a sibling's parents were also the respondent's parents, etc.)
2. If the reference person was the respondent's spouse or live-in partner, the roster member listed as "self" was actually a spouse or live-in partner, and the children relationship codes were salvaged.
3. If all the roster members other than the misidentified self were either roommates, boarders, or other nonrelatives, then the reference person was the respondent's roommate, boarder, or other nonrelative. All other relationship codes were salvaged.

8.2.5 Roster Edits for Other Household Members

Relationship codes were edited if the relationship of the roster member was impossible based on age and gender in relation to the self. Edits of roster ages, genders, and/or relationship codes were done that either changed the reported value to another value or changed the reported value to bad data. It is important to note that, in some cases, two members were selected in a household, which greatly increased the ability to edit the roster for those respondents. Some edits were triggered far less often in the 2002 NSDUH due to the introduction of consistency checks, as discussed in Section 8.2.2. Interviewer's explanations for overrides to consistency checks, and evaluations of their legitimacy, are given in Appendix I.

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

The following edits were performed on the roster age, gender, and relationship code values, 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 possible 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 still was believable even with the incorrectly entered age (e.g., "other relative"), this type of error was difficult to detect. On the other hand, if an age entered this way triggered one of the consistency checks discussed in Section 8.2.2.2, the interviewer had an opportunity to correct the entry error. On those occasions where the age did not trigger a consistency check, detection of the error was possible if two pair members were selected in the household 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 an *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 years old in each roster and comparing it with the screener roster, it became readily apparent whether and how a correction should have been 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. In previous survey years, the relationship code for grandchild (9) and grandparent (11) were commonly confused. With the introduction of new consistency checks (consistency checks #2 and #3 in Section 8.2.2.2), this did not occur in the 2002 NSDUH. However, the following edit, used in previous survey years, was maintained in case of overrides: if the age of the respondent was at least 20 years older than that of the roster member, but the roster member was identified as a grandparent, the relationship code was changed to grandchild. Conversely, if the age of the respondent was at least 20 years younger than that of the roster member, but the roster member was identified as grandchild, the relationship code was changed to grandparent.

8.2.5.2 Edits to Relationship Codes: Changes to Missing Codes

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent, and no internally consistent value could have been used to replace it. These edits were performed after the edits listed in Section 8.2.5.1 were completed. The relationship code in these instances, as listed below, was set to a bad data code.

1. More than one roster member aged 15 years or older was listed as living together with the respondent as though married, or as being the respondent's spouse. (An edit for multiple spouses of the same gender was covered in a consistency check. Any overrides of consistency checks were individually checked, although no overrides of this check occurred in the 2002 NSDUH.) For all roster members with such relationship codes and ages, the relationship codes were set to missing.
2. A roster member aged 15 years or older was identified as a spouse, and another was listed as living together as though married. In this case, the spouse code was maintained and the partner code set to bad data. These cases were individually checked to make sure that it made sense to keep the spouse code and not the partner code.
3. The roster member was the respondent's parent, but was younger than the respondent. This should have been covered by consistency check #2 in Section 8.2.2.2, but overrides did occur. This edit was automatic for respondents under 15 years old, under the default assumption that the overrides were not legitimate; however, interviewer explanations for the overrides were still checked, so that the edit could

have been undone if necessary. (The two overrides of this type were not legitimate.) Overrides for respondents aged 15 or older were allowed to stand, although the interviewer explanations and the rosters in question were also checked to ensure the overrides were legitimate. (The single override of this type was legitimate.)

4. The roster member was the respondent's child, but was older than the respondent. This should have been covered by consistency check #3 in Section 8.2.2.2, but overrides did occur. This edit was automatic for respondents under 15 years old, under the default assumption that the overrides were not legitimate; however, interviewer explanations for the overrides were still checked, so that the edit could have been undone if necessary. (The single override of this type was not legitimate.) Overrides for respondents aged 15 or older were allowed to stand, although the interviewer explanations and the rosters in question were also checked to ensure the overrides were legitimate. (The single override of this type was legitimate.)
5. The roster member was the respondent's biological parent, but was fewer than 12 years older than the respondent.
6. The roster member was the respondent's biological mother, but was more than 60 years older than the respondent.
7. The roster member was the respondent's biological child, but was fewer than 12 years younger than the respondent.
8. 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.
9. The roster member was the respondent's parent-in-law or child-in-law, but either the roster member or the respondent was under 15 years old.
10. The roster member was the respondent's child-in-law, but was at least 10 years older than the respondent.
11. The roster member was the respondent's parent-in-law, but was at least 10 years younger than the respondent.
12. 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 years, the code was set to child rather than to missing
13. The respondent had two children-in-law, but no children in the household. The in-law codes were set to missing.
14. The roster member was the respondent's grandchild, but the respondent was 25 years old or younger. If the "grandchild" was older than the respondent, it was covered by a consistency check (consistency check #3 in Section 8.2.2.2); no overrides to this consistency check occurred in the 2002 NSDUH.
15. 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, and the respondent was less than 24 years older than the roster member. As with the previous edit, if the "grandchild" was in fact older than the respondent, this was covered by consistency check #3 in Section 8.2.2.2, for which no overrides occurred in the 2002 NSDUH.

16. The roster member was the respondent's sibling and the previous roster member was a parent, but the roster member's age was within 4 years of the age of the parent. If the sibling was a half- or step-sibling, an additional requirement was that there was only one parent.
17. 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. If the roster member was a "grandchild" who was older than the respondent, then this was covered by consistency check #3 in Section 8.2.2.2. Similarly, if the roster member was a "grandparent" who was younger than the respondent, then this was covered by consistency check #2 in Section 8.2.2.2. In the latter two cases, the edit was automatic under the default assumption that the overrides were not legitimate; however, interviewer explanations for the overrides were still checked, so that the edit could have been undone if necessary. (The single override of the "grandparent" consistency check in the 2002 NSDUH data was not legitimate.)
18. If the respondent had two parents, but both parents were listed as biological mothers or biological fathers, the roster genders of both roster members were set to missing.

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

In Section 8.2.5.2, nonsensical relationship codes were set to bad data. Often, this occurred because the interviewer used someone other than the respondent as the reference person for one or more roster members. In some of these cases, the structure of the roster could have been used to determine the appropriate relationship code for that individual. Scenarios where the nonsensical code was "child" are listed below.

1. The interviewer might have put a roster member after the respondent's parent in the household roster. If the relationship code for that roster member was given as "child," the relationship code was nonsensical if the age made it impossible for the roster member to have been the respondent's child. (See #7 in Section 8.2.5.2.) In fact, if more than one "child" was listed after the respondent's parent, each would have been listed as 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 old, the relationship code was changed to sibling.
2. In some cases, the interviewer's entry for a roster member listed as child might have simply been a typographical error, where the "3" should have been a "4." Interviewers usually corrected such errors when a consistency check was triggered in cases where the child was older than the parent, or the child was a biological child who was less than 12 years younger than the parent. However, in cases where the interviewer insisted on the code, or where the child was younger than the respondent, but was less than 12 years younger than the respondent and was not biological, these typographical errors were more difficult to detect. If the respondent was living with parent(s), 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.

3. Both sides in a selected pair were respondents 18 or under, 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. If the age of the roster member was under 25, the relationship code was changed to sibling.
4. A roster member was listed as the respondent's child, who was not more than 12 years younger than the respondent, and the respondent was 25 or younger. The previous roster member was listed as grandparent. The "child" was in reference to the respondent's grandparent and should be considered either the respondent's parent or the respondent's uncle/aunt. If the roster member's age was at least 12 years older than the respondent and there were no non-immediate family codes (7, 12, 13, or 14 as described in Exhibit 8.2) then no uncles/aunts lived in the household. If a pair was selected, no non-immediate family codes were found in either pair member's roster. In either case, the relationship code was set to parent. Otherwise, one could not have been sure, so the relationship code was set to missing.

8.2.5.4 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. This most commonly occurred when a selected child had a parent with a spouse (the other parent) or live-in partner ("living together as though married"). Rather than identifying this individual as a "parent" or "other nonrelative," the interviewer identified the roster member as a spouse or live-in partner of the child, even though they intended for the point of reference to 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 identified a spouse/live-in partner, but were not part of a spouse-spouse pair. This only legitimately could have occurred if there were multiple spouse-spouse pairs in the household. In this edit, an attempt was made to identify cases with a single spouse-spouse pair in the household, where one pair member had a correctly identified spouse/live-in partner, and the other pair member's spouse/live-in partner was incorrectly identified. If the younger respondent, who was 25 years old or younger and at least 10 years younger than the older respondent, indicated a parent and the older respondent indicated neither parents nor parents-in-law, the older respondent should have been considered either the younger respondent's parent or the parent's spouse/partner. If the misidentified code was "spouse," the code was then changed to "parent." However, if the misidentified code was "live-in partner," the roster member may or may not have been considered the parent of the respondent. In most cases where the misidentified live-in partner was the respondent's parent's live-in partner, the code was then changed to parent. The exception occurred when (1) the live-in partner of this respondent's parent was the other respondent selected in a pair, and (2) the live-in partner did not indicate that the other pair member selected was his or her child in the parenting experiences question, FIPE3. In this instance, the

relationship code was changed to a special code indicating that the roster member was a live-in partner of the respondent's parent.

2. As in the previous edit, both sides in a selected pair identified a spouse/live-in partner, but were not part of a spouse-spouse pair, and there was only a single spouse-spouse pair in the household. In this edit, however, both sides incorrectly identified the spouse/live-in partner. In most cases, the pair was a sibling-sibling pair. If both respondents were under 21, both indicated a parent in the household, and the age difference between the respondents and their respective "spouse/live-in partner" was unusually large, then on each side the misidentified spouse/partner should have been considered a spouse/partner of the respondent's parent. If the misidentified codes were both "spouse," the codes were then changed to "parent." As stated above, however, if the misidentified codes were both "live-in partner," it was not clear whether each misidentified code should be "parent" or not. The rules used to determine whether the roster member was the respondent's parent were the same as in the previous edit (#1). The same special code as in the previous edit was used to identify a live-in partner of the respondent's parent.

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

4. Only one side in a selected pair identified a live-in partner, but the live-in partner was identified even though either (1) the respondent was under 15 or (2) the live-in partner was under 15. If the respondent listed one parent, but the other pair member listed two parents, the pair was a sibling-sibling pair, and the relationship code was in reference to the parent's live-in partner. The relationship code was changed to parent. If the respondent listed one fewer sibling than the other pair member, and the age difference between the respondent and the roster member identified as live-in partner was less than 15 years, the pair was a sibling-sibling pair, and the live-in partner code was changed to sibling.

5. Both sides in a pair identified the same household member as spouse or live-in partner. If the previous roster member on one of the sides was a sibling, the spouse/live-in partner should have been considered the sibling's spouse/live-in partner. The spouse relationship code was changed to "other relative." The live-in partner relationship code was changed to "other non-relative." If both sides had a previous roster member who was a sibling, it was not clear to which pair member the spouse/live-in partner belonged. To maintain proper counts, the spouse/live-in partner code for the youngest pair member was changed.

6. A roster member was identified as a spouse or live-in partner who was actually the spouse or live-in partner of the respondent's parent, but information from the other pair member was not used. As in previous edits, only one parent was identified (so

that the "spouse" or "live-in partner" could potentially have been the other parent), and this parent was the roster member listed before the "spouse/live-in partner." The additional rules for identifying this situation differed slightly for households with only one respondent selected and for paired respondents. For singly selected respondents in a household, (1) either the respondent was under 17 years old or the respondent was between 17 and 20 years old and the "spouse/live-in partner" was older than the respondent's parent; and (2) the respondent was more than 15 years younger than the "spouse/live-in partner." In the case of the misidentified spouse, the "spouse" of the respondent was considered the respondent's other parent. In the case of the misidentified live-in partner, the "partner" of the respondent was considered the live-in partner of the respondent's parent. Here, too, the code was changed to "parent." For members of a pair, the additional rules for the "spouse" were that (1) the respondent was under 21, or the "spouse" was older than the respondent's parent; and (2) the difference in age between the "spouse" and the respondent was greater than 12 years. As with the singly selected respondents, the relationship code was changed to be the respondent's other parent. The additional rules for the "live-in partner" were that (1) the respondent was under 15, or the "live-in partner" was older than the respondent's parent; and (2) the difference in age between the "live-in partner" and the respondent exceeded 5 years. If the "live-in partner" was the other pair member selected, and he or she did not identify the respondent as his or her child in the FIPE3 question, then a code was affixed that identified the roster member as a live-in partner of the respondent's parent. Otherwise, the roster member was identified as the respondent's second parent.

7. In all cases where the respondent was under 15 years old, he or she identified a spouse/live-in partner, and the above edits did not apply, the relationship code was set to bad data. In all but one case where the roster member was under 15, was identified as a spouse/live-in partner, and the above edits did not apply, the relationship code and roster age were set to bad data.[91]

8.2.5.5 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 the respondent's child or another sibling with the same characteristics, and either the respondent did not have parents in the household or the parent was a mother and the age difference between the mother and the "sibling" exceeded 50 years, 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.

[91] The single case that this edit missed was due to a programming error, which has been corrected for future years.

8.2.5.6 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 was possible that the roster member was a grandchild of a previous roster member. If two young respondents were selected where both identified the same grandparents and the same parents, and the respondent on the other side had siblings, the grandchild should have been considered the respondent's sibling. However, if this was not established, the roster member could have been the respondent's sibling or the respondent's cousin, so the code was set to bad data. If the "grandchild" was older than the respondent, it would have been covered by a consistency check (consistency check #3 in Section 8.2.2.2); no overrides to this consistency check occurred in the 2002 NSDUH.

8.2.5.7 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 or herself as the parent-in-law of the roster member. An in-law code was deemed invalid if a roster member was listed as the respondent's child-in-law, who was not more than 12 years younger than the respondent, and the respondent was 25 or younger. If the relationship code was listed as child-in-law, and the previous roster member was listed as grandparent, then the "child-in-law" was in reference to the respondent's grandparent and should have been considered either the respondent's parent or the respondent's uncle/aunt. If the roster member's age was at least 12 years older than the respondent and there were no non-immediate family codes (7, 12, 13, or 14 as described in Exhibit 8.2) then no uncles/aunts lived in the household. If a pair was selected, no non-immediate family codes were found in either pair member's roster. In either case, the relationship code was set to parent. Otherwise, there was no certainty associated with the relationship code, so this code was set to missing.

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), ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only), PRNTYP1-PRNTYP25 (type of parent: biological, adoptive, etc.), SIBTYP1-SIBTYP25 (type of sibling: biological, adoptive, etc.), CHDTYP1-CHDTYP25 (type of child: biological, adoptive, etc.), TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither).

8.4 Creation of Household Roster-Derived Variables

After replacing faulty information in the roster with missing values, the number of individuals with various characteristics in each roster was determined. These counts were recorded in the household roster-derived variables shown in Exhibit 8.3. If any information in the roster was missing, the roster-derived variable was set to missing. However, if some of the roster records for a respondent's household had missing data, then roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could have been imputed. Details on the imputation of the household roster-derived variables are given in Section 8.5. If two respondents were selected in a single household as part of a pair, the information from one pair member was not used to edit that of the other pair member. This was due to the fact that the interviews for each pair member could have occurred at different times, resulting in possible differences in the household composition.

Exhibit 8.3 Household Roster-Derived Variables

Variable Description	Variable Name
Total number of rostered people	TOTPEOP
Number of people in household aged 17 or younger	KID17
Number of people in household aged 65 or older	HH65
Indicator of whether the respondent had family members in household (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 younger than or equal to 17 years old	NRCH0_17
Number of respondent's children in household 18 to 20 years old	NROLDRCH
Number of respondent's children in household 21 or older	NROLDCH
Number of roommates/housemates in household	NROOMATE
Indicator of presence of mother in household (12 to 17 year olds)[1]	IMOTHER
Indicator of presence of father in household (12 to 17 year olds)[1]	IFATHER

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

The respondent's household size was assumed to equal the total number of rostered people in the household, TOTPEOP, as shown in Exhibit 8.3. The value of TOTPEOP was expected to equal the value of QD54 in most cases. However, in some cases the assigned self did not match, even approximately, the respondent's age or gender, or 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 the value of 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 members had missing ages, the values of KID17 and HH65 were missing, as well, regardless of whether some of the roster members were eligible to 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 was set to missing. However, if one of the nonmissing roster member's relationship codes indicated that the household contained one of the respondent's family members, the value of FAMSKIP was not missing even if other roster members had missing relationship codes.

Ten 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 8.3. Each of these variables was missing if the age or relationship codes for at least one roster member in a respondent's household were missing.

8.5 Imputation of Household Roster-Derived Variables

Although 14 roster-derived variables were created from the edited roster, missing values were imputed for only 4 of these variables: TOTPEOP, KID17, HH65, and FAMSKIP. The missing values in these variables were imputed using the univariate predictive mean neighborhood (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 were 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 were used as covariates for subsequent variables, if needed. The order in which the roster variables were imputed is shown in Exhibit 8.4.

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

Roster Variable	Edited Variable	Imputed Variable
Total number of rostered people	TOTPEOP	IRHHSIZE
Total number of children 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 as described in Exhibit 8.2. FAMSKIP was set to 1 if no relationship codes were missing, and the roster had codes of 1, 7, 12, and/or 14 as described in Exhibit 8.2.

8.5.2 Setup for Model Building

Once the hierarchy of the roster-derived variables was established, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all roster-derived variables were conducted separately within the four age groups: 12 to 17 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. (In the 2002 NSDUH, the final analysis weights were used if they were available. Because the final weight adjustments were completed at the time of the roster imputations, the final analysis weights were used.[92]) Item respondents were not defined across all roster categories; hence, this adjustment was computed separately for each age group and for each variable. The covariates in the response propensity models were the same covariates as those used in the main model considered in the next section. The item response propensity model is a special case of the generalized exponential model (GEM).[93] Greater details of the GEM software are presented 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 the LOGLINK procedure in SUDAAN® software.[94] The binary variable FAMSKIP was modeled using weighted logistic regression. The covariates in each model were continuous centered age,[95] continuous centered age squared, gender, race/ethnicity, imputation-revised roster-derived variables earlier in the sequence, region, population density, percentage Hispanic households in segment, percentage of owner-occupied households in segment, and (for TOTPEOP only) number of people in the household eligible for interviewing (from the pre-interview screener). There were also predictors

[92] In subsequent text, the use of the word "weights" will refer to the final analysis weights.

[93] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

[94] SAS®-callable SUDAAN® was used to fit the binary logistic regression models. Details about the LOGLINK procedure are discussed and additional references are provided in the *SUDAAN® User's Manual, Release 8.0* (RTI, 2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

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

that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups, the additional covariates of marital status, education status, and employment status were also included.

8.5.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods

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

8.5.5 Assignment of Imputed Values

Separate assignments were performed within each of the four age groups. A univariate imputation was implemented for each of the roster-derived variables within each age group, using the predictive means from the appropriate models. Assignments were made within preset bounds, as discussed in the next section. If no imputed values were available within the preset bounds, a random imputation was performed within those bounds.

8.5.6 Constraints on Univariate Predictive Mean Neighborhoods

A univariate imputation was implemented on each variable within each age group after predictive means 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 on roster-derived variables already imputed. The assignment of imputed values for KID17 was restricted within a lower and upper bound based on the value of IRHHSIZE and the nonmissing ages in the roster. For example, if a household roster had four members, with two 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. Likewise, HH65 was restricted within bounds in the same manner, using the variables IRHHSIZE and 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 have been considered a likeness constraint, which could have been loosened by enlarging delta, or abandoning the neighborhood altogether and taking the donor with the closest predictive 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 F.

9. Income Imputation

9.1 Introduction

As with most of the imputation-revised variables discussed in the previous chapters of this report, imputations for the 2002 National Survey on Drug Use and Health (NSDUH)[96] were accomplished using the predictive mean neighborhoods (PMN) technique, as described in Appendix C. The edits applied to the income variables are also described in this chapter.

The imputation of income was separated into two phases. The first phase was known as the "binary variable phase" and involved the imputation of all the binary income variables, as well as the number of months on welfare. This included the "yes-no" questions about the sources of income for the respondent and for the respondent's family living in the respondent's household, the number-of-months-on-welfare question (the only nonbinary variable in the binary variable phase), and a "yes-no" question regarding whether the respondent's income or the respondent's family income (in the household) was $20,000 or more (including income from the sources referenced in the previous questions). The correspondence between these questionnaire items and the edited variables is given in Exhibit 9.1. 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.

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

Exhibit 9.1 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[1]**	**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	FAMINC1
Total Income Specific Categories	QI21A, QI21B, QI23A, QI23B	PINC2	FINC2	FAMINC2

* Edited variables were not generated.

[1] Variables prefixed with "OFM" refer to all family members in the household other than the respondent. On the other hand, the variables FINC1 and FINC2 include information for all family members in the household including the respondent. In either case, if the respondent was the only family member in the household, as indicated by the family skip variable (IRFAMSKP = 1), these variables would have had legitimate skip codes. Moreover, a legitimate skip was assigned to the OFMxxx variable if the response to the personal income variable was "yes."

9.2 Edited Income Variables: Binary Variable Phase

9.2.1 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[97] From these two parts, three edited income 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 were 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 coded as a legitimate skip.[98] 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 sources of income and some of the other family sources of 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.

9.2.2 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 specifying 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 (QI20 and QI22), the second question in the pair applied to the entire family. In a similar manner to the source of income variables, the raw and edited versions of the personal total income questions (QI20 and PINC1, respectively) were nearly equivalent. The only case where equivalence did not occur was when the total family income question (QI22) was answered as "less than $20,000" and the total personal income question (QI20) was not answered, in which case PINC1 was logically assigned to be "less than $20,000." The second question in the pair asked about total family income, but was skipped if the respondent had no other family members in the household. The edited variable FINC1 was created by assigning legitimate skips in those cases. Moreover, if the total personal family income variable (QI20) was answered as "$20,000 or more" and the

[97] The computer-assisted interviewing (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 was 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.

[98] When the family skip variable IRFAMSKP indicated no other family members were 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 have been coded in the other-family-member variable would have overwritten real data, rather than a NSDUH blank data code. However, such cases rarely occurred.

total family income question (QI22) did not have a concurring answer, the value of FINC1 was logically assigned to be "$20,000 or more," regardless of the value of QI22.[99] A third binary total family income variable FAMINC1 was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household.

9.3 Imputed Income Variables: Binary Variable Phase

9.3.1 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. A motivation for using a hierarchy in PMN is given in Appendix C for drug use variables. For a model predicting whether a respondent had a given source of income, other sources of income were useful covariates. Following a provisional imputation of missing income values in the binary variable phase, the indicators earlier in the sequence were used as covariates for income models later in the sequence. Any imputed values in the income variables were considered temporary at this stage. This was 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 9.2.

9.3.2 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 as 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 versus $20,000 or more). Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. (As with health insurance, the final analysis weights were used as weights. See Section 10.3.2 for further discussion.) Because 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 the generalized exponential model (GEM), which is described in greater detail in Appendix B. The covariates were the same as those included in the main model, which is discussed in the next section.

9.3.3 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

[99] Logically, the 2002 NSDUH questionnaire should not have asked QI22 (total combined family income) if the answer to QI20 (total personal income) was "$20,000 or more." This problem will be fixed in the 2003 NSDUH questionnaire.

adjusted weights. For the models, the parameters were estimated using logistic regression.[100] 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 9.2. The covariates in each model were centered continuous age,[101] centered age squared, gender, race/ethnicity, provisional income indicators earlier in the sequence, region, population density, percentage Hispanic population, percentage 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 centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups, the additional covariates of marital status,[102] education status, and employment status were used. For the State rank groups, definitions were determined in terms of the proportion of a given State's residents having an income greater than or equal to $20,000.

Exhibit 9.2 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 used all of the predictors mentioned above except months on welfare.

[100] For the first time in the 2002 NSDUH, the logistic regression models were run in SAS®-callable SUDAAN® rather than SAS®. Both SAS® and SUDAAN® yield the same predictive means given the same set of covariates, but because SUDAAN® acknowledges the survey design, it gives correct values for the standard errors associated with each parameter estimate. Details about the logistic regression model and additional references can be found in the *SUDAAN® User's Manual, Release 8.0* (RTI, 2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

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

[102] In the 2002 NSDUH, for the first time, the marital status covariate was collapsed from four categories to three for the age group containing respondents aged 18 to 25. This was done in order to stabilize the regression models, thus producing more reliable predictive means. The instability was caused by the paucity of respondents in this age group who were widowed or divorced/separated; thus, these two categories were aggregated. The resulting three marital status categories were considered to be "currently married," "previously married," and "never married."

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,[103] such that $Y = \text{logit}(p)$ and p = number of months on welfare divided by 12. The binary total family income variable was modeled using weighted logistic regression. For a complete summary of the income imputation models, see Appendix E.

9.3.4 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 have been used as covariates in subsequent models. Although no provisional imputed values were used to build the models, it was necessary to calculate predictive means for all respondents, including item nonrespondents, using the parameter estimates from the models. This sometimes required the use of the provisional values for the covariates. The predicted probabilities from these models were used to assign provisional values using the univariate predictive mean neighborhood (UPMN) imputation method described in Appendix C.

9.3.5 Assignment of Provisional Imputed Values

Separate assignments of provisional values were performed within each of the four age groups (12 to 17 years, 18 to 25 years, 26 to 64 years, 65 years or older) 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 9.3.8.

9.3.6 Constraints on Univariate Predictive Mean Neighborhoods

After predictive mean values from the model had been determined, a univariate imputation was implemented on each variable within each age group. In general, the PMN is restricted by two types of constraints: (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 in the binary income variable imputations, donors were required to have the same value for the family skip variable (IRFAMSKP) as the recipient. The neighborhoods for the binary income indicators were restricted so that candidate donors and recipients would have been within the same age group (12 to 17 years, 18 to 25 years, 26 to 64 years, 65 years or older). Models were built separately within these four groups, so this likeness constraint was never loosened. A small delta could have also been considered a likeness constraint, which could have been loosened by enlarging delta, or abandoning the neighborhood altogether and taking the donor with the closest predictive mean. This was the only likeness constraint that could have been loosened with the binary income provisional imputations.

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

9.3.7 Multivariate Assignments

The predictive means were calculated with edited family income variables (see Exhibit 9.2) 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 9.3.

Exhibit 9.3 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

9.3.8 Multivariate Imputation

Sections 9.3.1 through 9.3.7 summarize the specifics of separating the set of binary income variables (in the 2002 NSDUH) into item respondents and item nonrespondents. These sections also describe model building, computation of predictive means, and the assignment of imputed values for these measures using a univariate predictive mean. In most cases, however, these univariate assignments were only provisional. The final imputed values for these income measures were obtained using neighborhoods built on a vector of predictive means using the multivariate predictive mean neighborhood (MPMN) technique as described in Appendix C. Consistent with the univariate imputations, the multivariate assignments were done separately within four age groups: 12 to 17 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 9.1. 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[104] 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.

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

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

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. Of the variables outlined in Exhibit 9.1, 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 predictive 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

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

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 F.

9.3.9 Binary Income Recode: GOVTPROG

The dichotomous recoded income variable GOVTPROG indicated whether the respondent participated in any government assistance programs. It was created from four imputation revised variables: family Supplemental Social Security income (IRFAMSSI), family food stamps (IRFSTAMP), family welfare payments (IRFAMPMT), and family welfare services (IRFAMSVC). Although a variety of recoded variables was created, but not discussed in this document, GOVTPROG is described here because it was used as a covariate in subsequent health insurance models (see Chapter 10 for details on the imputation of missing values in the health insurance variables).

9.4 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 from $0 up to $20,000 by increments of $1,000. By the same token, respondents who answered the binary total income question as $20,000 or more were asked to enter a specific category of income from $20,000 up to $50,000 by increments of $5,000. 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 question was for the individual respondent and the second question was for the entire family. As with 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.[105] The edited variable was created by assigning legitimate skips in those cases. A third specific category family-total-income variable was created, which was equal to the response to the second question in the pair if other family members were present in the household. On the other hand, 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.

[105] 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 have still routed the respondent to the family income question. However, in the CAI roster edits, the family relationship codes would have been set to bad data (see Chapter 8). It was possible that the family skip variable (IRFAMSKP) would have then been imputed to indicate that no other family members were present in the household. Hence, the legitimate skip coded in the family income variable would have overwritten real data rather than a NSDUH blank data code. However, such cases rarely occurred.

9.5 Imputed Income Variables: Specific Category Phase

9.5.1 Hierarchy of Income Variables

Three income variables resulted from editing the questions in the income-specific category phase (see Exhibit 9.1). These three variables were all considered simultaneously using a failure time model, which is described in greater detail in Section 9.5.3. Because only one model was fit, no hierarchy was required.

9.5.2 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 as 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. (As with health insurance and the binary income variables, the final analysis weights were used as weights. See Section 10.3.2 for further discussion.) The item response propensity model is a special case of the GEM, which is described in greater detail in Appendix B. The variables included in the model, which predicted the probability of item nonresponse, were the same as those included in the main model. Greater details are given in the next section.

9.5.3 Sequential Model Building

The specific categories of income were modeled using the LIFEREG procedure in SAS/STAT® software.[106] 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 NSDUH income intervals. The contribution of an individual with covariates in the matrix $\boldsymbol{X}$ to the overall likelihood is simply the probability mass assigned by the model to the interval $(l, u]$ containing the actual continuous income for that individual. For this interval, l represents the lower bound and u represents the upper bound. This contribution has the form $F(u|\boldsymbol{X},\beta,\sigma)$ -

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

$F(l|\boldsymbol{X},\beta,\sigma)$, where F is a cumulative distribution function. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.[107]

LIFEREG allowed several choices for the functional form of the parametric model that corresponded to the error distribution discussed earlier, including the two-parameter log-logistic, lognormal, gamma, and Weibull, and also the three-parameter generalized gamma. Each of these models was fit to each of the four age-group-specific datasets. Compared with the other models, the gamma distribution provided a better overall fit, as indicated by likelihood techniques. Because the three-parameter generalized gamma did not significantly improve on its two-parameter special cases, when using the likelihood ratio tests as criteria for comparison, it was decided to use a two-parameter model.

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

9.5.4 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 regression covariates. 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 C.

9.5.5 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

---

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

values using the same donor for all three variables. The multivariate imputation process is further described in Section 9.5.7.

9.5.6 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 F.

9.5.7 Multivariate Assignments

The predictive means were calculated using the edited (specific category) family income variables (see Exhibit 9.2) as the response variables. For each family income variable, neighborhoods were created using scalar-predictive means from the appropriate model. The methodology for determining the neighborhood was therefore univariate in terms of these scalar-predictive means. Three edited variables were associated with each predictive mean, so that the missing values for the three variables required assignment of imputed values. Hence, even when determining the provisional imputed values using the univariate procedure, the assignment of imputed values was multivariate for all but two of the variables. For the 2002 NSDUH, the imputation-revised variable for the personal income variable was called IRPINC2, the family income variable with legitimate skips was called IRFINC2, and the family income variable without legitimate skips was called IRFAMIN2.

9.5.8 Specific Category Income Recode: INCOME

The recoded variable INCOME classified the families of respondents into four income levels: less than $20,000; from $20,000 to $49,999, from $50,000 to $74,999; and greater than or equal to $75,000. It was a recode of the variable IRFAMIN2. A variety of recoded variables were created but are not discussed in this document; however, as with GOVTPROG, the variable INCOME is discussed here because it was used as a covariate in subsequent health insurance models (see Chapter 10 for details on the imputation of missing values in the health insurance variables).

10. Health Insurance

10.1 Introduction

Two methods were used to create the final imputation-revised health insurance variables. The first method, referred to as the "old method," followed the general strategy used in previous iterations of the National Survey on Drug Use and Health (NSDUH).[108] Specifically, this method was implemented to create two general imputation-revised health insurance variables. The first variable was simply an imputation-revised version of the edited private health insurance variable. For the second variable, a recoded overall health insurance variable was created by combining information from the edited health insurance variables; then, missing values for that recoded health insurance variable were imputed. Because the health insurance questions in the NSDUH changed every year between the 1999 survey and the 2001 survey, different versions of the overall health insurance variable were created for each of these surveys, two of which could be and were created using the questions available in the 2002 survey. Thus, a total of three imputation-revised health insurance variables were created from the 2002 NSDUH using the old method.

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

10.2 Edited Insurance Variables

Exhibit 10.1 shows the edited counterparts for some of the health insurance questionnaire (raw) variables. In the 2002 NSDUH, the edited variables had the same values as the questionnaire variables, except that missing values were replaced by standard NSDUH missing value codes.

10.2.1 Edited Insurance Variables (Old Method)

In the old method, three health insurance indicators were created from these six variables. Two of them, INSUR and INSUR3, indicated whether the respondent had any health insurance; the third, PINSUR, indicated whether the respondent had any private health insurance. INSUR3, which was consistent with the variable of the same name created in the 2001 NHSDA, was coded as "yes" if any one of the six variables listed in Exhibit 10.1 were coded as "yes," and "no"

---

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

if all six variables were coded as "no." The other overall insurance indicator, INSUR, was created to maintain consistency with the 1999 NHSDA. Because the questions associated with CHIPCOV and HLTINNOS 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 10.1 were coded as "yes," and "no" if all four variables were coded as "no." In the 2000 NHSDA, the variable INSUR2 was created to take advantage of the additional information provided by questions that did not exist on the 1999 questionnaire. However, because these additional questions were either replaced or reworded in the 2001 and 2002 surveys, the variable INSUR2 was not created in either of those surveys.

Exhibit 10.1 Mapping of Raw Health Insurance Variables to Edited Counterparts

Question Variable	Question Text	Edited Counterpart[1]
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)
QHI02A	Is the respondent currently covered by a Children's Health Insurance Program operated by your state of residence?[2] (Asked only of respondents aged 12 to 19)	CHIPCOV (1 = yes, 2 = no)
QHI03	Is the respondent currently covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)
QHI11	Is the respondent currently covered by any kind of health insurance, that is, any policy or program that provides or pays for medical care?	HLTINNOS (1 = yes, 2 = no, 99 = legitimate skip[3])

[1]Missing values in these edited values were represented by standard missing value codes.
[2]The questionnaire did not ask the question exactly in this way. It identified the specific program, depending upon the state of residence entered by the respondent.
[3]A respondent was assigned a legitimate skip for HLTINNOS if they answered "yes" or gave no answer to at least one of the other health insurance questions.

To create the variable for private health insurance, PINSUR, only the edited variable PRVHLTIN was used. Missing data for the edited variable PRVHLTIN were coded using the standard NSDUH missing data codes for "don't know," refused, and blank, whereas missing data for PINSUR were all coded as "98," which was a code for missing data. Except for the codes used to handle missing data, PINSUR and PRVHLTIN were equivalent. The variable PINSUR was created to maintain consistency with pre-1999 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.

10.2.2 Edited Insurance Variables (Constituent Variables Method)

In the constituent variables method, the editing process combined the variables MEDICAID and CHIPCOV to create the variable CAIDCHIP. This variable was the one that was later imputed to indicate whether someone was covered by Medicaid or one of the state children's health plans. All the other edited variables in Exhibit 10.1, except HLTINNOS, were used directly as base variables for imputation.

A respondent was routed to QHI11 if they answered no to all the other health insurance questions. All other respondents were given a legitimate skip value to the variable HLTINNOS, as shown in Exhibit 10.1. It was possible, therefore, that the imputation-revised versions of the four specific health insurance variables would all have had a value of "no," and the value of HLTINNOS would have been a legitimate skip, if one or more of the "no" values was imputed. In this instance, another variable was needed to reflect the fact that a respondent could have had a valid yes/no imputed value for "any other health insurance" even though the respondent was never asked QHI11, and HLTINNOS = "99," which was a legitimate skip code. This variable, which was called ANYOTHER, was created using HLTINNOS and an additional edited variable SKHLCCOV, which indicated whether a respondent was covered by any health insurance. SKHLCCOV was defined as follows:

SKHLCCOV = 1 (or 3) if CAIDCHIP=1, MEDICARE=1, CHAMPUS=1 or PRVHLTIN=1[109]

= 2 if CAIDCHIP=2, MEDICARE=2, CHAMPUS=2, and PRVHLTIN=2

= missing value code if the nonmissing values of CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN are all "2," and at least one of these variables had a missing response

ANYOTHER was therefore defined as follows:

ANYOTHER = legitimate skip code (99) if SKHLCCOV = 1 or 3

= SKHLCCOV if SKHLCCOV = 2 or a missing value code

10.3 Imputed Health Insurance Variables (Old Method)

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

---

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

10.3.1 Order of Modeling Health Insurance Variables (Old Method)

A multivariate predictive mean neighborhood (MPMN) imputation method for private health insurance and overall health insurance was implemented. However, respondents who answered "yes" to the private health insurance question were logically also covered by overall health insurance. Therefore, it was not possible to use INSUR or INSUR3 as covariates in the PINSUR model, or vice versa.

10.3.2 Setup for Model Building (Old Method)

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

In the 2002 NSDUH, one model was created for PINSUR and another for INSUR3. A respondent was considered an item respondent for health insurance only if his or her status was known for both private health insurance and overall health insurance as defined by INSUR3. To meet this criterion, the respondent must have had a valid "yes" or "no" response in PRVHLTIN (the edited variable corresponding to QHI06). In addition, he or she either must have answered QHI01, QHI02, QHI02A, QHI03, and QHI11 with a valid "no" response, or answered "yes" to at least one of the six questions (including QHI06). This ensured that the interview respondent's status with respect to both overall health insurance (INSUR3 definition) and private health insurance was completely known. For example, if the interview respondent did not answer QHI01, but answered "no" to the other five questions, his or her status with respect to overall health insurance depended on the missing response to QHI01. However, if the respondent answered "yes" to any of the other five questions, the value of INSUR3 was already known to be "yes."

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

To ensure that the weights adequately represented the population, the weights for item nonrespondents (as defined by INSUR3) were reallocated to item respondents using item response propensity models within each age group for the pair INSUR3 and PINSUR. (In the 2002 NSDUH, the final analysis weights were used in imputation procedures, if they were available. Because the final weight adjustments were completed at the time of the insurance imputations, the final analysis weights were used.[110]) The item response propensity model is a

[110] In subsequent text, the use of the word "weights" will in fact refer to the final analysis weights.

special case of the generalized exponential model (GEM).[111] Greater details of the GEM software 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.

10.3.3 Sequential Model Building (Old Method)

The probability that the respondent had health insurance (as defined by INSUR3) and the probability that the respondent had private health insurance were both modeled for item respondents, within each age group, using the nonresponse adjusted weights. The private health insurance model was created only for respondents who were known to have overall health insurance, so that the predicted probability modeled was *P*(PINSUR=1 | INSUR3=1). For the models, the parameters were estimated using logistic regression.[112] Each response propensity model included the following pool of predictors: continuous centered age,[113] race/ethnicity, centered age squared, centered age cubed, gender, population density, percentage of housing in segment that was owner-occupied, percentage of Hispanics in the segment, percentage of non-Hispanic blacks in the segment, and household size. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups (i.e., 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older), the additional predictors of marital status, education level, and employment status were also considered in each model.

10.3.4 Computation of Predictive Means (Old Method)

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

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

The final imputed values for overall health insurance (using both the INSUR and INSUR3 definitions) and private health insurance were obtained using neighborhoods built upon

---

[111] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

[112] In the 2002 NSDUH, the software used for most imputation modeling was SUDAAN®, whereas SAS® had been used in previous survey years. However, the logistic model for the old method of imputing health insurance variables used SAS® to maintain consistency with the practice of previous survey years. SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

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

a vector of predictive means. The vector had two elements: *P*(overall health insurance, as defined by INSUR3) and *P*(private health insurance | overall health insurance, as defined by INSUR3). For both overall and private health insurance, the imputation method used was the MPMN procedure. More details regarding this imputation method are presented in Appendix C. Similar to the response propensity models, the multivariate assignments were done separately within the same four age groups: 12 to 17 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 PINSUR, INSUR, and INSUR3 and was within the same age group. Logical constraints were placed on individuals who were missing one or two of the three indicators. Respondents who were missing either overall indicator, but did not have private health insurance, required donors who also did not have private health insurance.[114] If a respondent was only missing INSUR3, then INSUR must have been "no" because a "yes" value for INSUR would have necessarily meant that INSUR3 would have been "yes" and therefore nonmissing. Hence, donors must also have had a "no" value for INSUR. By the same token, if a respondent was only missing INSUR or was missing both PINSUR and INSUR, but not INSUR3, then INSUR3 must have been "yes" because a "no" value for INSUR3 would have necessarily meant that INSUR would have been "no" and therefore nonmissing. In this case, donors must also have had a "yes" value for INSUR3. Finally, respondents who indicated that they had health insurance but were missing the private health insurance indicator required donors who had some health insurance.[115] 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 had all applicable elements of the multivariate predictive mean vector "close to" (i.e., within the delta distance) the recipient's elements of the predictive mean vector. Because the imputation was multivariate, the set of deltas was also multivariate, where a different delta corresponded to each element of the predictive mean vector. Likeness constraints were loosened in the order 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 G. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in Appendix F.

The full predictive mean vector contained elements for overall health insurance (as defined by INSUR3) and private health insurance (conditional on a "yes" response to the overall health insurance (INSUR3) indicator). The portion of the full predictive mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If a respondent was missing INSUR, but not INSUR3, the predictive mean that was derived using INSUR3 was used. The portions of the full predictive

---

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

[115] Again, this technically was not a logical constraint. However, because all respondents who did not have health insurance also did not have private health insurance, and the recipient had health insurance, the distribution would have been skewed in favor of a "no" indicator if these respondents were allowed to be donors.

mean vector used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given in Appendix G. The Mahalanobis distance[116] 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 predictive means within a multivariate delta of the recipient's vector of predictive means, the neighborhood was abandoned, and the respondent with the closest Mahalanobis distance was selected as the donor. The procedure is described in detail in Appendix C.

10.4 Imputed Specific Health Insurance Variables (Constituent Variables Method First Stage)

The constituent variables method of creating the final imputation-revised health insurance variables amounted to imputing missing values in each of the edited health insurance variables that, when combined together, constituted "overall health insurance." In the first stage of this method, which is described in this section, four imputation-revised specific health insurance variables were created representing whether the respondent had health insurance from Medicaid or a state children's health insurance program (IRMCDCHP), Medicare (IRMEDICR), CHAMPUS (IRCHMPUS), or private health insurance (IRPRVHLT). Missing values in these variables were imputed in a multivariate imputation. These final variables were derived from the edited variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN, respectively. The second stage is described in Section 10.5.

10.4.1 Order of Modeling Health Insurance Variables (Constituent Variables Method First Stage)

The first step in imputing the four specific health insurance variables was to determine the order in which the variables would be modeled. A motivation for using a hierarchy in PMN for drug use variables is given in Appendix C. For a model predicting whether a respondent had a specific type of health insurance, other types of health insurance were useful covariates. Following a provisional imputation of missing health insurance values, the indicators earlier in the sequence were used as covariates for health insurance variables later in the sequence. Any imputed values in the health insurance variables were considered temporary at this point. This was due to the fact that the final imputation was not done for health insurance variables until the modeling was completed for all four specific health insurance variables. The order in which the health insurance indicators were imputed as follows: CAIDCHIP, MEDICARE, CHAMPUS, PRVHLTIN.

10.4.2 Setup for Model Building (Constituent Variables Method First Stage)

Once the hierarchy of health insurance variables was determined, the next step was to define respondents, nonrespondents, and the item response mechanism. For an individual to be considered an item respondent for the specific health insurance variables, he or she must have had complete data for the four edited specific health insurance variables. Imputation for CAIDCHIP, CHAMPUS, and private health insurance were conducted within the four age

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

groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. Imputation for Medicare was conducted within the following age groups: 12 to 17 year olds, 18 to 64 year olds, and respondents 65 years of age or older.[117]

Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. (In the 2002 NSDUH, the final analysis weights, appropriately poststratified and adjusted for unit nonresponse, were used.) The item response propensity model is a special case of the GEM. Greater details of the GEM software are presented in Appendix B. The covariates in the item response propensity model included a centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three level income variable. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the three older age groups (i.e., 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older), the additional predictors of marital status, education level, and employment status were also considered in each model.

10.4.3 Sequential Model Building (Constituent Variables Method First Stage)

Starting with CAIDCHIP, the probability that an individual was covered by a given type of health insurance was modeled for item respondents, within each age group, using the nonresponse adjusted weights. For the models, the parameters were estimated using logistic regression in SUDAAN®.[118] The predictors included in all models were centered age, centered age squared, gender, race/ethnicity, population density, and percentage of housing in that segment that was owner-occupied. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender except for the 65 years of age. For the three older age groups (i.e., 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older), the additional predictors of marital status, education level, and employment status were also considered in each model. Additional predictors were specific to each model, depending upon the response variable of interest, and are listed below.

CAIDCHIP: household size; a four-level family income variable;[119] binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, or social security; and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.

---

[117] The age groups 18 to 25 and 26 to 64 were combined for the Medicare variable because (1) only a small proportion of respondents in these age groups had Medicare, particularly for the 18 to 25 age group and (2) a respondent of working age could have only received Medicare if he or she was not working due to disability. This was true regardless of whether the respondent was 18 to 25 or 26 to 64 years old.

[118] SAS®-callable SUDAAN® was used to fit the binomial and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in the *SUDAAN® User's Manual, Release 8.0* (RTI, 2001). SAS® software is a registered trademark of SAS Institute, Inc.; SUDAAN® is a registered trademark of RTI International.

[119] The four levels of the family income variable were: under $20,000; $20,000 to $49,999; $50,000 to $74,999, and $75,000 or more.

MEDICARE: for respondents 18 or over, a binary indicator of whether the respondent was on social security; and for respondents under 18, a binary indicator of whether anyone in the respondent's family in the household received social security.

CHAMPUS: a binary indicator of whether the respondent (or the respondent's family in the household, if the respondent was under 18) received income from sources other than those given in the binary income questions (see Chapter 9 for details); a three-level income variable;[120] and for respondents 18 or over, an indicator of whether the respondent had ever been in the military service, designated by an imputation-revised version of the edited variable SERVICE.[121]

PRVHLTIN: household size; a four-level family income variable (the same variable that was used in the CAIDCHIP model); binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, social security, or from sources other than those given in the binary income questions (see Chapter 9 for details); and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.[122]

The complete summary of the health insurance models can be found in Appendix E.

10.4.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods (Constituent Variables Method First Stage)

Following the modeling for the four specific health insurance variables corresponding to CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN, in the sequence given in Section 10.4.1, missing values were replaced by provisional imputed values. This was necessary so that these variables could have been used as covariates in subsequent models. Although no provisional imputed values were used to build the models, it was necessary to calculate predictive means for all respondents, including item nonrespondents, using the parameter estimates from the models. This sometimes required the use of the provisional values for the covariates. The predicted probabilities from these models were used to assign provisional values using the univariate predictive mean neighborhood (UPMN) imputation method as described in Appendix C.

---

[120] The three levels were: under \$20,000; \$20,000 to \$49,000, and \$50,000 or over.

[121] The variable SERVICE generally had a very low level of missingness (1 missing value in the 2002 NSDUH). Since covariates in these models must not have had any missing values, the missing value in the SERVICE variable was randomly imputed as a "yes" if the random number was greater than the mean value of SERVICE across all the other respondents, and "no" otherwise.

[122] If the respondent did not have other family members in the household, the family income binary indicators listed as predictors were equivalent to the personal income binary indicators.

10.4.5 Assignment of Provisional Imputed Values (Constituent Variables Method First Stage)

Separate assignments of provisional values were performed within the age groups that were used for each of the respective first three health insurance variables.

10.4.6 Multivariate Imputation of the Specific Health Insurance Variables (Constituent Variables Method First Stage)

The final imputed values for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN were obtained using neighborhoods built upon a vector of predictive means. For these four variables, the imputation method used was the PMN procedure as described in Appendix C. Similar to the response propensity models, the multivariate assignments were done separately within the same four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. No logical constraints were applied to the health insurance variables, since no internal inconsistencies would have resulted from any type of donor. However, a number of likeness constraints were applied, depending upon the missingness pattern. The variables that were included as likeness constraints were highly correlated with the response variables, but (in most cases) could not have been included as predictors in the models due to the large number of missing values in the predictors. In general, any nonmissing values that the recipient had for CAIDCHIP, MEDICARE, CHAMPUS, or PRVHLTIN had to match between donor and recipient, though this constraint was often the first to be loosened. In addition, the donor's predicted mean(s) for each variable that was missing was required to be within 5 percent of the recipient's predicted mean(s). This was usually the last constraint to be loosened. Finally, specific likeness constraints were associated with each of CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Constraints associated with each variable are discussed briefly below. The order in which the constraints were loosened depended upon the missingness pattern, and is described in detail in Appendix F. The portions of the full predictive mean vector used to create the multivariate neighborhoods for each missingness pattern, with accompanying adjustments, are given in Appendix G.

CAIDCHIP

The donor and recipient had to have the same status regarding whether or not a respondent's family had received any government public assistance. This was measured by the variable GOVTPROG, which is described in detail in Section 9.3.9.

MEDICARE

A respondent of working age (between the ages of 18 and 64) could have only received Medicare if he or she were not working due to disability. If MEDICARE was missing, a constraint was included that required donors and recipients to have had the same status in this regard, using the appropriate level of the variable JBSTATR. This constraint was never loosened. In addition, the donor and recipient had to have the same status regarding whether or not a respondent's family had received social security.

CHAMPUS

In the models for CHAMPUS, two variables were included as covariates that were also used as likeness constraints. An imputation-revised version of the variable SERVICE (whether the respondent had ever been in the military service) was used in the CHAMPUS model, whereas SERVICE was used directly as a likeness constraint. The other variable was a binary indicator of whether the respondent (or the respondent's family in the household, if the respondent was under 18) received income from sources other than those given in the binary income questions (see Chapter 9 for details). Neither likeness constraint was loosened in the 2002 NSDUH for any of the age groups, making their inclusion in the models unnecessary.

PRVHLTIN

In the model for PRVHLTIN, a four-level income variable was used as a covariate that was also used as a likeness constraint for the youngest three age groups. This likeness constraint was never loosened in the 2002 NSDUH, making its inclusion in the models unnecessary for these three age groups. If it had been loosened, the donor and recipient would have been required to have the same value for a two-level income variable (under $20,000 and $20,000 or over). For respondents 65 years of age or over, this two-level income variable was used as an initial likeness constraint, and was never loosened in the 2002 NSDUH.

10.5 Imputed Any Other Health Insurance and Overall Health Insurance Recoded Variable (Constituent Variables Method Second Stage)

The constituent variables method of creating the final imputation-revised health insurance variables amounted to imputing missing values in each of the edited health insurance variables that, when combined together, constituted "overall health insurance." In the second stage of this method, which is described in this section, a variable is created (IROTHHLT) that indicates whether respondents had any type of health insurance, even though they reported or were imputed to have none of the four types of specific health insurance, as recorded by IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. The final overall health insurance indicator is created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT.

10.5.1 Order of Modeling Health Insurance Variables (Constituent Variables Method Second Stage)

Only one variable required imputation in the second stage. An order of imputation was therefore unnecessary.

10.5.2 Setup for Model Building (Constituent Variables Method Second Stage)

Imputation for the any other health insurance variable was conducted within the following age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents 26 years of age or older.[123] For a respondent to be considered an item respondent for modeling the any other health

[123] Three age groups were used instead of four due to the small number of respondents who would have been included in the 65+ age group.

insurance variable, he or she first had to be part of the domain, which included respondents who had either a reported or imputed "no" value to all four imputation-revised specific health insurance variables (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT). Among respondents who were part of the domain, item respondents had to have complete data for the variable ANYOTHER, as defined in Section 10.2.2. Response propensity adjustments were computed within each age group in order to make the item respondent weights representative of the entire domain. (In the 2002 NSDUH, the final analysis weights, appropriately poststratified and adjusted for unit nonresponse, were used.) The item response propensity model is a special case of the GEM. Greater details of the GEM software are presented in Appendix B. The covariates in the item response propensity model included a centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three-level income variable. There were also predictors that consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the two older age groups (i.e., 18 to 25 year olds, and respondents 26 years of age or older), the additional predictors of marital status, education level, and employment status were also considered in each model.

10.5.3 Sequential Model Building (Constituent Variables Method Second Stage)

The probability that an individual was covered by any other health insurance was modeled for item respondents within the domain defined in the previous section, within each age group, using the nonresponse adjusted weights. The parameters were estimated using logistic regression in SUDAAN®, with the same base set of predictors that were used for the specific health insurance variables. In particular, these included centered age, centered age squared, gender, race/ethnicity, population density, percentage of housing in that segment that was owner-occupied, and a three level income variable. This base set also consisted of one-way interactions of centered age with race/ethnicity, centered age with gender, race/ethnicity with gender, centered age squared with race/ethnicity, and centered age squared with gender. For the two older age groups (i.e., 18 to 25 year olds and respondents 26 years of age or older), the additional predictors of marital status, education level, and employment status were also considered in each model. Additional predictors were specific to the any other health insurance model: household size, binary indicators of whether the respondent's family in the household received income from public assistance, wages, interest, Social Security, and for respondents 18 or older, a binary indicator of whether the respondent had other family members in the household.[124]

The complete summary of the health insurance models can be found in Appendix E.

10.5.4 Computation of Predictive Means and Univariate Predictive Mean Neighborhoods (Constituent Variables Method Second Stage)

Following the modeling of the any-other-health-insurance variable, missing values were replaced by imputed values. In the usual way, predictive means were calculated for all respondents, including item nonrespondents, using the parameter estimates from the models. The

[124] If the respondent did not have other family members in the household, the family income binary indicators listed as predictors were equivalent to the personal income binary indicators.

predicted probabilities from these models were used to assign imputed values using the UPMN imputation method as described in Appendix C.

10.5.5 Assignment of Imputed Values (Constituent Variables Method Second Stage)

Separate assignments of provisional values were performed within the three age groups. The imputed values from these assignments were considered final. The imputation-revised version of the any other health insurance variable was called IROTHHLT.

10.6 Creation of the Final Overall Health Insurance Variable (Constituent Variables Method)

The final overall health insurance indicator was created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT. If a respondent had a reported or imputed "yes" value for any of these five variables, the respondent was considered to have health insurance. Otherwise, he or she did not have health insurance. This was recorded using the variable IRINSUR4, to be distinguished from the overall health insurance variable that was created using the old method, IRINSUR3. Though IRINSUR4 was technically a recoded variable created from other variables, an imputation indicator was nevertheless created, called IIINSUR4. Specifically, IIINSUR4 was set to "3" if any of the five constituent health insurance variables were imputed, "2" if none of the five variables were imputed and at least one was logically assigned, and "1" otherwise.

References

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

Bowman, K. R., Chromy, J. R., Odom, D. M., & Penne, M. A. (2003, April). The 2001 National Household Survey on Drug Abuse: Sample design report. In *2001 National Household Survey on Drug Abuse: Methodological resource book* (Section 2, prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 10; RTI/07190.330.400). Research Triangle Park, NC: RTI.

Chen, P., Emrich, S., Gordek, H., Penne, M. A., Singh, A. C., & Westlake, M. (2002, July). Person-level sampling weight calibration for the 2000 NHSDA. In *2000 National Household Survey on Drug Abuse: Methodological resource book* (Section 3, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 10, RTI/7190.374.100). Research Triangle Park, NC: RTI.

Chromy, James R. (1979). Sequential sample selection methods. In *Proceedings of the American Statistical Association, Survey Research Methods Section*. Washington, DC: American Statistical Association.

Cox, B. G., & Cohen, S. B. (1985). *Methodological issues for health care surveys*. New York: Marcel Dekker, Inc.

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 721-726). Washington, DC: American Statistical Association.

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 estimating in survey sampling. *Journal of the American Statistical Association, 87*, 376-382.

Draper, N. R., & Smith, H. (1981). *Applied regression analysis* (2nd ed.). New York: John Wiley & Sons.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 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.

Iannacchione, V. (1982). Weighted sequential hot deck imputation macros. In *Proceedings of the Seventh Annual SAS Users Group International Conference*. Cary, NC: SAS Corporation.

Kroutil, L. A. (2003a, June). General principles and procedures for editing drug use data in the 2001 NHSDA computer-assisted interview. In *2001 National Household Survey on Drug Abuse: Methodological resource book* (Section 14, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 28, RTI/7190.395). Research Triangle Park, NC: RTI.

Kroutil, L. A. (2003c, June). Procedures for editing interviewer-administered data in the 2001 NHSDA computer-assisted interview. In *2001 National Household Survey on Drug Abuse: Methodological resource book* (Section 14, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 28, RTI/7190.395). Research Triangle Park, NC: RTI.

Kroutil, L. A. (2003b, June). Procedures for editing supplementary self-administered data in the 2001 NHSDA computer-assisted interview. In *2001 National Household Survey on Drug Abuse: Methodological resource book* (Section 14, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 28, RTI/7190.395). 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.

Office of Applied Studies. (2001). *Development of computer-assisted interviewing procedures for the National Household Survey on Drug Abuse*. Retrieved January 28, 2004, from http://www.samhsa.gov/oas/nhsda/CompAssistInterview/toc.htm

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register, 62* (210), 58781-58790. [Available at http://www.gpoaccess.gov/fr/index.html]

Penne, M. A., Lessler, J. T., Bieler, G., & Caspar, R. (1998). Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI field experiment. In *Proceedings of the American Statistical Association, Social Statistics Section* (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 Triangle Park, NC: Author.

RTI. (2003, February). 2002 National Survey on Drug Use and Health: Content of 2002 NSDUH instrument. In *2002 National Survey on Drug Use and Health methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-98-9008, Deliverable No. 28, RTI 7190). Research Triangle Park, NC: RTI.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics, 4*(1), 87-94.

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). 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.

Williams, R. L., & Chromy, J. R. (1980). SAS sample selection MACROS. In *Proceedings of the Fifth International SAS Users Group International Conference* (pp. 382-396). Cary, NC: SAS Corporation.

Appendix A: Hot-Deck Method of Imputation

Appendix A: Hot-Deck Method of Imputation

A.1 Introduction

Typically, with the hot-deck method of imputation, missing responses for a particular variable (called the "base variable" in this appendix) are replaced by values from similar respondents with respect to a number of covariates (called "auxiliary variables" in this appendix). If "similarity" is defined in terms of a single predicted value from a model, these covariates can be represented by that value. The respondent with the missing value for the base variable is called the "recipient," and the respondent from whom values are borrowed to replace the missing value is called the "donor."

Although only two hot-deck imputation methods were used in the 2002 National Survey on Drug Use and Health (NSDUH), [125] three different methods are discussed in this document: unweighted sequential hot deck, unweighted random nearest neighbor hot deck (NNHD), and weighted sequential hot deck. The first method, the unweighted sequential hot deck, was the exclusive method of hot-deck imputation used for the 1991 to 1998 surveys and the paper-and-pencil interviewing (PAPI) sample of the 1999 survey. This method was used for all demographic variables in the 1999 survey, but no other variables. In the 2000 NHSDA, the unweighted sequential hot deck method was only used for education and employment status, and was not used at all in 2001 or 2002 surveys. However, it remains in this appendix for historical purposes and for the sake of comparison with the other two methods. In a similar manner to the 1999 (computer-assisted interviewing [CAI] sample of the survey), 2000, and 2001 surveys, the 2002 NSDUH primarily used the second hot-deck method listed, the unweighted random NNHD. The third hot-deck method, weighted sequential hot deck, incorporated the sampling weights associated with each respondent. Starting in the 2002 NSDUH, the immigrant variable imputations described in Chapter 5 utilized the weighted sequential hot-deck method. For more information on weighted sequential hot-deck imputation, see Cox (1980, pp. 721-725).

A step that is common to all hot-deck methods is the formation of imputation classes, which is discussed in Section A.2. This is followed by a general description of the three hot-deck methods Sections A.3-A.5. With each type of hot-deck imputation, the identities of the donors are generally tracked. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

A.2 Formation of Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation procedures were implemented independently within classes defined by the cross of the auxiliary variables. These classes were defined by logical and likeness constraints, which are described in the main body of this report. Classes defined by the likeness constraints were collapsed if

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

insufficient donors were available, and classes defined by logical constraints were not collapsed, due to the possibility of a resulting inconsistency with preexisting nonmissing values.

A.3 Unweighted Sequential Hot Deck

In the years that the unweighted sequential hot deck was used, its implementation involved three basic steps. After the imputation classes were formed, the file was appropriately sorted and imputed values were assigned as described in the following sections.

A.3.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous NSDUHs to sort the files prior to imputation:

- **Straight Sort**. A set of variables was sorted in ascending order by the first variable specified; then within each level of the first variable, the file was sorted in ascending order by the second variable specified; and so forth. For example:

1	1	1
1	1	2
1	2	1
1	2	2
1	3	1
1	3	2
2	1	1
2	1	2
2	2	1
2	2	2
2	3	1
2	3	2

- **Serpentine Sort**. A set of variables was sorted so that the direction of the sort (ascending or descending) changed each time the value of a variable changed. For example:

1	1	1
1	1	2
1	2	2
1	2	1
1	3	1
1	3	2
2	3	2
2	3	1

2	2	1
2	2	2
2	1	2
2	1	1

The serpentine sort has the advantage of minimizing the change in the entire set of auxiliary variables every time any one of the variables changes its value.

A.3.2 Replacing Missing Values

The file was sorted and then read sequentially. Each time an item respondent was encountered (i.e., the base variable was nonmissing), the base variable response was stored, updating the donor response. Any subsequent nonrespondent that was encountered received the stored donor response, creating the statistically imputed response. A starting value was needed if an item nonrespondent was the first record in a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value. Due to the fact that the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

A.3.3 Potential Problem

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there was the risk of several nonrespondents appearing next to one another on the sorted file. To detect this problem in the NSDUH, the imputation donor was identified for every item being imputed. Then, when frequencies by imputation donor were examined, the problem was detected if several nonrespondents were aligned next to one another in the sort. When this problem occurred, sort variables were added or eliminated, or the order of the variables was rearranged.

A.4 Unweighted Random Nearest Neighbor Hot Deck

As with the unweighted sequential hot deck, the unweighted random NNHD was implemented in three steps. After the imputation classes were formed, a neighborhood of potential donors was created, from which imputed values were assigned, as described in the following sections.

A.4.1 Creating a Neighborhood of Potential Donors

First, a metric was defined to measure the distance between units, based on the values of the covariates. Then a neighborhood was created of potential donors "close to" the recipient based on that metric. For example, the distance between the values of the recipient and potential donors for each of the auxiliary variables were calculated, then the donors for the neighborhood were chosen such that the maximum of these distances was less than a certain value, referred to as "delta." This neighborhood was restricted, using the imputation classes defined above, so that the potential donors' values of the base variable were consistent with the recipient's preexisting nonmissing values of related variables. In the NSDUH, the values of the auxiliary variables were represented by a predicted mean from a model, so that the distance metric was a univariate

Euclidean distance between the predicted mean of the recipient and the potential donors. The distance was relative when dividing this value by the predicted mean of the recipient, resulting in delta as a percentage.

A.4.2 Randomly Selecting a Donor for the Recipient from the Neighborhood of Donors

From the neighborhood of donors created in the previous step, a single donor was randomly selected. The base variable values for this single donor replaced those of the recipient. The selection was conducted as a simple random sample because weights were incorporated in determining the neighborhood mean, which was the predicted mean. Alternatively, a weighted selection could have been employed if weights had not been used to determine the neighborhood mean.

A.5 Weighted Sequential Hot Deck

The steps taken to impute missing values in the weighted sequential hot deck were equivalent to those of the unweighted sequential hot deck. The details on the final imputation, however, differed with the incorporation of sampling weights. The first step, as always, was the formation of imputation classes. Afterwards, two additional steps, as described below, were implemented.

A.5.1 Sorting the File

Within each imputation class, the file was sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that were better predictors for the item being imputed were used as the first sorting variables). In general, two types of sorting procedures were used in previous NSDUHs to sort the files prior to imputation: straight sort and serpentine sort. Both of these methods are described in detail in Section A.2.2.

A.5.2 Replacing Missing Values

The procedure used in the 2002 NSDUH followed directly from Cox (1980). Specifically, once the imputation classes are formed, the data is divided into two data sets: one for respondent and one for nonrespondents. Scaled weights $v(j)$ are then derived for all nonrespondents using the following formula:

$$v(j) = w(j)s(+)/w(+); \; j = 1, 2, \ldots n$$

where n is the number of nonrespondents, $w(j)$ is the sample weight for the j^{th} nonrespondent, $w(+)$ is the sum of the sample weights for the all nonrespondents, and $s(+)$ is the sum of the sample weights for all the respondents (Cox, 1980). The respondent data file is partitioned into zones of width $v(j)$, where the imputed value for the j_{th} nonrespondent is selected from a respondent in the corresponding zone of the respondent data file.

This selection algorithm is an adaptation of Chromy's (1979) sequential sample selection method, which could be implemented using the Chromy-Williams sample selection software

(Williams & Chromy, 1980). Furthermore, Iannacchione (1982) revised the Chromy-Williams sample selection software, so that each step of the weighted sequential hot deck is executed in one macro run.

A.5.3 Benefits of Weighted Sequential Hot-Deck

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there is the risk of several nonrespondents appearing next to one another in the sorted file. An imputed value could still be found for those cases, since the algorithm would select the previous respondent in the file; however, some modifications are required in the sorting procedure to prevent a single respondent from being the donor for several nonrespondents (see Section A.3.3). With the weighted sequential hot-deck method, on the other hand, this problem does not occur because the weighted hot deck controls the number of times a donor can be selected. In addition, the weighted hot deck allows each respondent the chance to be a donor since a respondent is selected within each $v(j)$.

The most important benefit of the weighted sequential hot-deck method, however, is the elimination of bias in the estimates of means and totals. This type of bias is particularly present when the response rate is low or the covariates explain only a small amount of variation in the specified variable. In addition, many surveys sample subpopulations at different rates, and using the sample weights allows, in expectation, the imputed data for the nonrespondents to have the same mean (for the specified variables) as the respondents. In other words, the weighted hot deck preserves the respondent's weighted distribution in the imputed data (Cox, 1980).

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

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

B.1 Introduction

For the 2002 National Survey on Drug Use and Health (NSDUH),[126] as well as previous surveys, a special case of the generalized exponential model (GEM) was used for weighting procedures. This special case was known as the item response propensity model, where weights among item respondents were adjusted to account for the weights of the item nonrespondents. The GEM macro, which was written in SAS/IML® software,[127] was developed at RTI International for weighting procedures. Additional technical details concerning the GEM are given in the following sections.

B.2 Distance Function

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

$$\Delta(w,d) = \sum_{k \in s} \frac{d_k}{A_k} \left\{ (a_k - 1_k) \log \frac{a_k - 1_k}{c_k - 1_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\} \qquad \text{(B2.1)}$$

where $a_k = w_k / d_k, A_k = (u_k - 1_k)/(u_k - c_k)(c_k - 1_k)$, and λ_k, c_k, u_k are prescribed real numbers. Let T_x denote the p-vector of control totals corresponding to predictor variables (x_1, ..., x_p). Then the calibration constraints for the above minimization problem are

$$\sum_{k \in s} x_k d_k a_k = T_x \qquad \text{(B2.2)}$$

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

$$a_k(\lambda) = \frac{\lambda_k (u_k - c_k) + u_k (c_k - \lambda_k) \exp\{A_k x_k' \lambda\}}{(u_k - c_k) + (c_k - \lambda_k) \exp\{A_k x_k' \lambda\}} \qquad \text{(B2.3)}$$

Note that the number of parameters in GEM should be $\leq n$, where n is the size of the sample s. This is also the dimension of vectors d and w. It follows from (B2.3) that

---

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

[127] SAS® software is a registered trademark of SAS Institute, Incorporated.

$$\ell_k < a_k < u_k, k = 1,..,n \qquad \text{(B2.4)}$$

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

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

and $a_k(\lambda) = \exp(x_k{}'\lambda)$.

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

B.3 GEM Adjustments for Extreme Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters $\lambda_k, c_k,$ 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 b_k = med$\{d_k\} \pm 3*$ IQR if d_k is an outlier, where IQR represents the interquartile range and is a measure of dispersion for a data set, and the quartiles for the weights are defined with respect to a suitable design-based stratum. Then with GEM for outlier treatment, $\lambda_k = 1, c_k = c = 1 + \sum_{s^{**}} (d_k - b_k) / \sum_{S^*} d_k$ 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^{**} denotes 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, λ_2 is set as $\lambda_2 = 1, c_2 = \rho^{-1}, u_2 = u > \rho^{-1}$, where ρ is the overall response propensity. For outliers with high weights, ℓ_k is set as $\ell_k = \ell_1 m_k$. In addition, $c_k = m_k$ and $u_k = u_1 m_k$, where $m_k = b_k / d_k$, and $\lambda_1 < 1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, $1 < \lambda_3 < \rho^{-1} = c_3 < u_3$ is set for outliers with low weights.

For the poststratification (ps) adjustment, λ_k is set for nonoutliers as $\lambda_k = \lambda_2, c_k = c_2 = 1, u_k = u_2$, 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, there exists the flexibility of specifying different bounds for different subsamples, as well as making the lower bound (in the case of outlier and nr adjustments) 1 by choosing the center $c_k > 1$.

B.4 Newton-Raphson Steps

Let X denote the n x p matrix of predictor values, and for the ν^{th} iteration,

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

where

$$\Phi_k^{(\nu)} = (u_k - a_k^{(\nu)})(a_k^{(\nu)} - l_k)/(u_k - c_k)(c_k - l_k).$$

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

$$\lambda^{(\nu)} = \lambda^{(\nu-1)} + (X'\Gamma_{\Phi,\nu-1} X)^{-1}(T_x - \hat{T}_x^{(\nu-1)}) \quad \text{(B4.1)}$$

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

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

B.5 Scaled Constrained Exponential Model

In previous NSDUHs, constrained exponential models (CEMs) were used for ps and scaled CEMs were used 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., $\lambda_k = \lambda, u_k = u$, and $c_k = c = 1$ such that $\lambda < 1 < u$). 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 λ in CEM as ρ, it ensures that the scaled adjustment factor for nonresponse is at least 1.

Appendix C: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

Appendix C: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

C.1 Introduction

The 2002 National Survey on Drug Use and Health (NSDUH)[128] used a predictive mean neighborhood (PMN) method for imputing missing values. This method was implemented in the past several surveys. Starting with the 1999 survey, this PMN method was a new approach, which was developed for the imputation of missing values in the computer-assisted interviewing (CAI) sample. This approach has been used since the 1999 NHSDA[129] and can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, PMN 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 PMN method is a combination of two commonly used imputation methods: a nonmodel-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 NSDUH, the set of covariates typically included demographic variables, as well as some other nonmissing drug use variables. In the case of the NSDUH, to further ensure that a donor matched the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with drug use, such as age categories, were often used to restrict the set of donors. Furthermore, other restrictions involving outcome variables were imposed on the neighborhood.

Note that in NNHD, unlike sequential hot deck, a distance function is used to define closeness between the recipient and a donor. So, there is less of a problem of sparseness of the donor class, but the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

---

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

[129] In the surveys after the 1999 one, only a CAI sample was selected.

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 the recipient's predictive mean (which takes continuous values) and the donor's outcome value (which takes discrete values) is not well defined.

C.2.2 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

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

The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be tedious in practice because of the computational problems due to the volume of model parameters and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, none of the existing solutions takes 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 1999–2002 surveys, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for the use of univariate techniques for fitting multivariate (but conditional) 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 all the missing values in an incomplete record is desirable because this method preserves the relationships among the variables of interest. Moreover, if the predictive mean vector includes both missing and nonmissing variables (this could easily happen when models are fitted in a univariate manner under a hierarchy), it is also ensured 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. This is due to the fact 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 in the recipient record.

C.3 Outline and Description of Method

The procedure for implementing UPMN and MPMN in the 2002 NSDUH entailed six steps. Steps 2 through 5, and sometimes Step 6, were cycled through each of the drugs and drug use measures in the order determined by Step 1. Steps 4 and 5 (Steps 4 through 6 when applicable) could have been considered a variant of a random NNHD.

C.3.1 Step 1: Definition of Hierarchy

The first step was to determine the order in which variables were modeled, so that variables early in the hierarchy could have been used for modeling the conditional predictive mean (i.e., they have the potential to have been part of the set of covariates for variables later in the hierarchy). Note that usually not all variables in the hierarchy were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For example, in the drug modules in the 2002 NSDUH, different drugs needed to have been modeled, with different measures of drug use for each drug. It was therefore necessary to determine the order in which the combination of drugs and drug use measures would have been handled. Using the sequence of variables determined by this step, the procedure involved cycling through Steps 2 through 5, and sometimes Step 6. In the application of the PMN to the NSDUH, the order of imputation for drugs was determined by considering such factors as the level of stigma associated with the drugs, the level of "missingness" in the data (see Appendix G), and the degree to which one set of drugs could have been used as predictors for other drugs. The order of drugs was given by cigarettes, smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, crack, and heroin. The order of drug use measures imputed was determined based on the natural hierarchy of the variables: lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and age of first use.

For each variable, Steps 2 through 5 were followed for the NSDUH.

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

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

C.3.3 Step 3: Sequential Hierarchical Modeling

The model was built using the complete data respondents only with weights adjusted for item nonresponse. For the drug modules in the 2002 NSDUH, lifetime usage indicators were modeled first because all other drug use indicators depended on an indication of lifetime use or nonuse. Once the hierarchy of drugs for lifetime usage was determined, lifetime usage indicators for individual drugs were modeled in a sequential fashion. The sequence used for the remaining combinations of drugs and drug use measures depended on what covariates were desired in the models and what variables were considered part of a multivariate set.

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

Once the model was fitted, the predictive means for item respondents and item nonrespondents were calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element out of that vector was chosen, so that each respondent had exactly one predictive mean value.[130] This predictive mean was the matching variable in a random NNHD. It could have come directly from the model, it could have been adjusted to account for the conditioning on the time period, or (if it was the predicted value based on a model with a transformed response variable) it could have been back-transformed to the original units.

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

The pool of donors was further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a "logical constraint," was given by age at first crack use, which must not have been less than age at first cocaine use. Other constraints, called "likeness constraints," were placed on the pool of donors to make the attributes of the neighborhood as close to that of

[130] Alternatively, a provisional MPMN method could have been performed by using the predicted probabilities from the polytomous model. Consequently, the final MPMN would have been built based on probabilities from the polytomous model, as well as predictive means for the other variables in the multivariate set. See Step 6 (Section C.3.6) for a description of the MPMN.

the recipient as possible. For example, for age at first use, the age of the donor and the age of the recipient were restricted to have been the same whenever possible, and the donor and recipient must have come from States with similar usage patterns. A small value of delta could have also been considered as a likeness constraint. Whenever insufficient donors were available to meet the likeness constraints, including the preset small value of delta, the constraints were loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors was available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor was found using a sequential hot deck, where matching was done on the predictive mean. (Even though weights would not have been used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process because weighting would already have been incorporated in the calculation of the predicted mean.)

If many variables were imputed in a single multivariate imputation, it was advantageous to preserve, as much as possible, correlations between variables in the data. However, the more variables that were included in a multivariate set, the less likely that a neighborhood could have been used for the imputation within a given delta. Even though there were many advantages to using multivariate imputation, one disadvantage, in several instances, was not being able to find a neighborhood within the specified delta.

C.3.5 Step 5: Assignment of Imputed Values Using a Univariate Predictive Mean Neighborhood

Using a simple random draw from the neighborhood developed in Step 4, a donor was chosen for each item nonrespondent. If only one response variable was imputed, the assignment step was a simple replacement of a missing value by the value of the donor. It was possible, however, that a donated quantity was a function of the final imputed value. For example, for 12-month frequency of drug use, because donors and recipients could potentially have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of the total period was donated rather than the observed 12-month frequency, where the "total period" could have ranged up to a year. In the assignment step, the donor's proportion of total period was multiplied by the recipient's maximum possible number of days in the year that he or she could have used the substance.

The assignment step was multivariate if several response variables were associated with a single predictive mean, provided more than one of those response variables was missing. In that case, all of the missing values were imputed using the same donor. If there was more than one response variable associated with a single predictive mean, but not all of them were missing, only the missing values were replaced by those of the donor. The resulting imputed values were provisional if a multivariate predictive mean vector was needed in a final multivariate imputation; otherwise, these values were final.[131]

[131] If the variable was part of a multivariate set upon which the MPMN method was applied, and provisional values were not needed for subsequent models, Steps 4 (creation of delta neighborhood) and 5 could have been skipped.

The variables requiring imputation were part of a multivariate set if a multivariate predictive mean vector was used to match donors and recipients in a final multivariate imputation. If the variables were part of a multivariate set, it was necessary to cycle through Steps 2 through 5 for each variable in the set, then proceed to Step 6 after completing Steps 2 through 5 for the last variable in the set. If the variables were not part of a multivariate set, then it was only necessary to go through Steps 2 through 5 once, and proceeding to Step 6 was unnecessary. After the completion of either Step 5 (if a univariate predicted mean was used) or Step 6 (if a multivariate predictive mean vector was used), the next variable in the hierarchy requiring imputation was processed by returning to Step 2.

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

With the MPMN method, the neighborhood was defined based on a vector of predictive means rather than from a single predictive mean as in the univariate case. This vector may have encompassed 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 was calculated between the elements of this vector of predictive means, where the observed values were missing, and the corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predictive means that were calculated on the basis of past year and past month users were further adjusted to account for the probability that a respondent was a past year user or a past month user. For example, in the 2002 NSDUH, the full predictive mean vector for alcohol included the following elements:

1. *recency, past month*: *P*(past month alcohol user | lifetime alcohol user);

2. *recency, past year, not past month*: *P*(past year but not past month alcohol user | lifetime alcohol user);

3. *12-month frequency*: *P*(the respondent used alcohol on a given day in the past year | past year user of alcohol) * *P*(past year user of alcohol | lifetime alcohol user);[132]

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

[132] For the 12-month frequency, 30-day frequency, and 30-day binge frequency, the models were fitted using logits. These logits were converted to probabilities when creating the predictive mean vector. Interpreting the proportion of the year used as a probability of use on a given day in the year assumed that the probability of use on each day in the year was equal. This, of course, was not true. However, the violation of this assumption did not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it did allow a predicted mean to be made conditional on what was known.

5. *30-day binge frequency*: *P*(the respondent was a binge drinker on a given day in the past month | past month user) * *P*(past month alcohol user | lifetime alcohol user).

The subset of elements used to determine a neighborhood for a particular item nonrespondent depended on the missingness pattern of that item nonrespondent.[133] Moreover, if partial information was available on the recency of use, the predictive means was adjusted to account for that knowledge. For example, if a particular item nonrespondent was known as a past year alcohol user and his 12-month frequency was known, the elements above for which differences would have been calculated would be element #1 conditioned on past year use, and elements #4 and #5. That is,

P(Past month alcohol user | Lifetime alcohol user) ÷ *P*(Past year alcohol user | Lifetime alcohol user),

P(Respondent used alcohol on a given day in the past month | Past month user of alcohol)* *P*(Past month alcohol user | Lifetime alcohol user) ÷ *P*(Past year alcohol user | Lifetime alcohol user), and

P(Respondent was a binge drinker on a given day in the past month | Past month user) * *P*(Past month alcohol user | Lifetime alcohol user) ÷ *P*(Past year alcohol user | Lifetime alcohol user).

A neighborhood that resulted from this vector of distances was constrained by a multivariate preset delta, where the distances associated with each element of the predictive mean vector must each have been less than the preset delta associated with that element. From the donors that remained, a single neighborhood was 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 was only calculated for those respondents who met the delta constraint. The neighborhood was determined by selecting the *k* smallest Mahalanobis distances within this subset of item respondents for a given item nonrespondent.

For those variables in the response vector that were not missing, only those that were missing were replaced. However, logical constraints must have been placed on the multivariate neighborhood, so that imputed values were consistent with preexisting nonmissing values. For example, if a respondent was missing a 30-day frequency, but his or her nonmissing 12-month

[133] Alternatively, the entire predictive mean vector could have been used 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 was accomplished using UPMN, which is computationally much faster.

frequency was 350, a donor could not have had a 30-day frequency smaller than 350 - 335, or 15. If the number of respondents in the univariate subset who met the logical constraints, imposed upon the multivariate neighborhood, was fewer than *k* but greater than 0, all the respondents in the resulting subset were selected for the neighborhood. Finally, if there were no respondents within the univariate subset who met the logical constraints imposed by the multivariate neighborhood, the *k* smallest Mahalanobis distances who met the logical constraints among all candidate donors for a given item nonrespondent were selected for the neighborhood. In addition to the multivariate delta, likeness constraints were used to make the donors in the neighborhood as much like the recipient as possible. These could have been loosened if insufficient donors were available. Finally, as with the univariate neighborhood, an unweighted sequential hot deck was used as a last resort if there were not enough sufficient donors available who met the logical constraints and the loosest set of likeness constraints allowable.

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

C.4 Comparison of PMN with Other Available Imputation Methods

The PMN methodology addresses all of the shortcomings of the unweighted sequential hot-deck method:

- **Ability to use covariates to determine donors is far greater than in the hot deck.** As with other model-based techniques, using models allows more covariates to be incorporated, including measures of use of other drugs, in a systematic fashion, where weights can be incorporated without difficulty. However, like a hot deck, covariates not explicitly modeled can be used to restrict the set of donors using logical constraints. If there is particular interest in having donors and recipients with similar values of certain covariates, they can be used to restrict the set of donors using likeness constraints even if they are already in the model

- **Relative importance of covariates is determined by standard estimating equation techniques.** In other words, there are objective criteria based on methodology, such as regression, 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.

- **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 be difficult to incorporate directly in the model.

- **Sampling weights are easily incorporated in the models.** The weighted hot deck can be viewed as a special case of PMN.

- **Correlations across response variables are justified by making the imputation multivariate.**

- **Choice of donor can be made random by choosing delta large enough 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 with mean zero can be mimicked. This helps to avoid bias in estimating means and totals, variances of which can be estimated as in two-phase sampling or by suitable resampling methods.

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

For the 2002 National Survey on Drug Use and Health (NSDUH),[134] it was not uncommon for a respondent to have felt that the categories for race or Hispanicity given in the questionnaire did not apply to him or her. In these situations, interviewers were given the opportunity to manually enter (type) a category that the respondent felt best described himself or herself. The manually entered responses were called "other-specify" or "alpha-specify" responses because they were typed in a part of the question that asked the interviewer to specify an alphabetic response. These alpha-specify responses were then matched to codes to describe the responses, which were collected and maintained in a file known as a "dictionary." Other-specify responses from each survey year were matched against this file, and any responses without codes were given new codes and added to the dictionary; therefore, the size of the dictionary file increased each survey year. (In most cases, new unmatched responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican.") 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 the 2002 questionnaire, three core questions (QD05, QD05ASIA, and QD06) focused on the respondent's race. Respondents were permitted to select more than one race in QD05. If they selected "Asian" as one of their races, they were routed to QD05ASIA, where they were also permitted to select more than one answer. There also was a follow-up question (QD06) asking respondents who selected multiple races in QD05 and/or QD05ASIA to select among those chosen the single race that best described them. Respondents had the opportunity to direct the interviewer to select "other" as the race in both QD05 and (if applicable) QD05ASIA, whereby the interviewer then typed the alphabetic response given by the respondent. The alpha-specify responses to these two questions were considered simultaneously. The only instance where separate codes were required for the two questions occurred when the interviewer marked the Asian category, then manually entered "Indian" as the alphabetic response. Normally, "Indian" would have mapped to a code for American Indian, but in this case the respondent would have been considered Asian Indian. The race questions used in the 2002 survey are as follows:

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

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

1 White
2 Black/African American
3 American Indian or Alaska Native (American Indian includes North American, Central American, and South American Indians)
4 Native Hawaiian
5 Other Pacific Islander
6 Asian (for example: Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese)
7 Other (Specify)

QD05ASIA: Which of these Asian groups best describes you? Just give me the number or numbers from the card.

1 Asian Indian
2 Chinese
3 Filipino
4 Japanese
5 Korean
6 Vietnamese
7 Other (Specify)

QD06: Which **one** of these groups, that is [races chosen in QD05 and QD05ASIA], **best** describes you?

1 White
2 Black/African American
3 American Indian or Alaska Native (American Indian includes North American, Central American, and South American Indians)
4 Native Hawaiian
5 Other Pacific Islander
6 Asian Indian
7 Chinese
8 Japanese
9 Filipino
10 Korean
11 Vietnamese
12 [Other from QD05ASIA, if applicable]
13 [Other from QD05, if applicable]
14 None of these

D.2.1 Race Alpha Responses

The other-specify responses were examined when (a) "other" was selected as a race in the race questions (QD05 and/or QD05ASIA),[135] 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).[136] In such cases, if a valid other-specify response was given, the code corresponding to that response was used in order to assign a value of EDRACE, the base variable for imputing IRRACE, and NEWRACE, the base variable for imputing IRNWRACE (see Chapter 4). In many cases, the interviewers entered an alpha-specify response that could have been mapped directly to 1 of the 12 categories in the race questions. Otherwise, other codes were used for which various algorithms were used to determine the final racial category. The codes could have been classified into general categories, which are described below:

1. The following other-specify responses and their derivatives were classified as "black/African American": Afro American, Haitian, Caribbean Creole, African or any country from sub-Saharan Africa except Namibia or South Africa (see #6 below), negra or negro, St. Vincent. Also, any respondent indicating that he or she was part Hispanic (or "Spanish") and part black was classified as black.

2. The following responses and their derivatives were considered within the "Asian/Pacific Islander" group for EDRACE, but were given separate codes for NEWRACE: Native Hawaiian, Other Pacific Islander (which also included Micronesian, Polynesian, Samoan, Saipan, and Guamanian), Chinese (which also included Taiwanese, Cantonese, Guanma), Filipino, Japanese, Asian Indian (which also included Nepalese, Pakistani, Bengali [Bangladesh], Hindu, Indian American, African Indian, Kashmirian, Punjabi, Sri Lankan, Sikh), Korean, Vietnamese, Other Asian. The "Other Asian" group included the following responses and their derivatives: Lao, Thai, Cambodian, Kampuchean, Malaysian, Burmese, Myanmar, Okinawan, Chaldean, East Indian, Indonesian, Eurasian, Iranian, Persian, Kurd, Afghan, Hmong, Kazakh, Mienh, Singaporean, Mongolian, Tibetan, Uzbek, Turkmenistan. A separate code was also given to cases indicating "Asian" with no specific group.

3. The following other-specify responses and their derivatives were classified as "American Indian": American Indian or Alaska Native: Native American, Indian (except respondents who also indicated they were Asian), Indigenous, Mayan, Aztec, Yaqui, Zapotec, Apache, Blackfoot, Cherokee, Navajo, Tewa,

[135] Although it was a possibility that a respondent could have given conflicting other-specify races in QD05 and QD05ASIA, this did not occur in the 2002 survey.

[136] There were a few exceptions to this rule. The alpha-specify answer to QD04 (QD04RACE) was also selected over the alpha-specify answer to QD05 (QD05RACE) if QD04RACE had more specific race information than QD05RACE that did not contradict that given in QD05RACE.

Weott, Aleut, and Eskimo. Also, any respondent indicating that he or she was part Hispanic (or "Spanish") and part American Indian was classified as "American Indian."

4. The following other-specify responses and their derivatives were classified as "white": Caucasian, north African or any country from north Africa, Arabic, Turkish, Armenian, Jewish, Middle Eastern/Israeli, Assyrian, any country from central, eastern, or southeastern Europe except Germany, blanco, Celtic, Anglo-Saxon, Armenian, Cajun, Caledonian, any combination of European nationalities, or part-Hispanic and part-white. (A separate code was available for Middle Eastern countries, but they were all finally classified as white.)

5. If a respondent indicated a 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, Mexicans and Central or South Americans combined, Mexicans and Cubans combined, Puerto Ricans and Central or South Americans combined, Puerto Ricans and Cubans combined, Cubans and Central or South Americans combined, and Hispanic non-white (for example, trigueno="dark").

6. For certain countries of origin given in the other-specify responses, race was randomly assigned using census data for those countries. Canada was added to this list for the 2002 survey. 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:[137]

- 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;

[137] Note that these percentages were used to randomly assign respondents to races although the distribution of assigned races in the sample did not match these percentages exactly. Also note that if 0 percent was statistically imputed, no respondents were assigned to those races.

- 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;
- Australia: 95 percent white, 4 percent Asian, 1 percent black, 0 percent statistically imputed;
- Barbados: 80 percent black, 16 percent mixed, 4 percent white; and
- Canada: 87 percent white, 9.9 percent Asian, 1.6 percent black, and 1.5 percent American Indian.

7. If the respondent indicated a mixture of races in the alpha-specify responses, the particular mixture was recorded with a separate code. For example, a respondent who answered "black and white" was given the code 202, while a "Korean and Chinese" respondent was given the code 310. Respondents who said "mestizo” or "mestiza" were classified as "American Indian and white" and given the code 203. Respondents with these codes involving at least one non-Asian were classified into the more than one race category in NEWRACE, while respondents with more than one race code involving only Asians were classified as "Asian multiple categories" in NEWRACE. The EDRACE value assigned is described in the following section.

Mainly to prevent respondents from being incorrectly classified as multiple races, one or more of the alpha-specify responses were ignored if any of the following were true:

- The respondent indicated "Hispanic" in the QD05 alpha-specify response, but was already known to be Hispanic from QD03.
- One or more of the respondent's alpha-specify responses were technically not a race, and the respondent selected at least one of the listed race categories in QD05. The following alpha-specify answers were ignored by this rule: nonwhite non-specific, Hispanic nonwhite, Hispanic, any Hispanic-origin group or

combination of Hispanic-origin groups, "Spanish" or some variant, any country of origin listed in #6 above, and any combination of the above answers.

- One or more of the respondent's alpha-specify responses were redundant, in that it echoed what was already known from the selections from the listed race categories in QD05. For example, if the respondent selected "black/African American" in QD05, and gave "Haitian" as an alpha-specify answer, the alpha-specify response was ignored.

D.2.2 Assigning a Race When Multiple Races Were Selected (EDRACE/IRRACE Only)

As stated earlier, respondents were allowed to select more than one race when responding to QD05 or QD05ASIA, 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 was in place to determine among the multiple races chosen which one best described those 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 specifically for respondents who indicated more than one race, regardless of whether they indicated a single race in QD06.

D.2.3 Race Dictionary Codes

If a single response was given to the specific categories in QD05 and QD05ASIA, and no alpha-specify responses were given, a code between 1 and 12 was assigned based on this response. If more than one response was given but none was an alpha-specify response, the respondent was set aside and identified as "more than one race," "Asian multiple categories," or "Hawaiian and other Pacific Islander." Otherwise, a code was assigned based on the respondent's alpha-specify responses (codes 21 to 985). For the 2002 survey, codes 21 to 32 were equivalent to codes 1 to 12, except that the race identification was obtained from the alpha-specify responses. The values of EDRACE and NEWRACE were obtained using these codes (see Section D.2.2), as follows:

1	White
2	Black/African America
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 (includes Burmese/Burma)
30	Korean
31	Vietnamese
32	Other Asian (includes Iran, Kurd, Afghan, Chaldean, Laos, Cambodia, Kampuchea, Krum)
33	Asian nonspecific
34	Guamanian
35	Non-white non-specific
40	Hispanic non-white (incl. trigueno=”dark”, moreno, brown)
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
56	Barbados
57	West Indies
80	United Kingdom
81	Portugal/European Spanish
82	Spanish, maybe European
83	Other Western Europe (including Albania)
84	Middle East/Israel/North Africa
85	Canada
86	New Zealand
87	South Africa (Zambian, Namibia, Zimbabwe)
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
146	Cuban and Asian
147	Mexican and Other Pacific Islander
148	Puerto Rican and Other Pacific Islander
149	Central or South American and Other Pacific Islander
150	Cuban and Other Pacific Islander
151	Mexican & European Spanish
152	Puerto Rican & European Spanish
153	Cuban & European Spanish
154	Central or South American & European Spanish
155	Dominican & European
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
422	White, Asian, Hispanic
900	Mixed
901	Mezclado, Mezclada (Hispanic mixed)
985	Bad data
994	"Unknown"/"Don't Know"
997	"Rather Not Say"/"Refused"("American" or "All of Them")

D.3 Hispanicity

As with the race questions, Hispanic respondents[138] had the opportunity to specify a Hispanic-origin group by responding "other" to the Hispanic-origin group question (QD04). Also, respondents were permitted to select multiple Hispanic-origin groups in response to QD04. However, there was no follow-up question asking respondents to choose a single group from among multiple groups chosen. Below is the Hispanic-origin group question.

QD04: Which of these Hispanic, Latino, or Spanish groups best describes you? Just give me the number or numbers from the card.

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican
3 Central or South American
4 Cuban/Cuban American
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 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.

4. The following other-specify responses were classified as "Central or South American": Central or South American, and Central or South American countries, including countries that are not typically Hispanic (Belize, Guyana, etc.).

5. The following other-specify responses were classified as "Caribbean Islander": Hispanic Caribbean Islander (includes Dominican Republic and

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

Santo Domingo), Dominican (where Dominica vs. Dominican Republic unclear), "Other Caribbean."

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, a 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 QD04 (codes 1 through 4), or on their Hispanicity alpha-specify responses (codes 11 through 85). For the 2002 survey, codes 11 through 14 were equivalent to codes 1 through 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.3.1), which are presented below. Values 1 through 4 come directly from the questionnaire responses; values 11 through 14 come from the alpha-specify responses.

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
27	Central or South American/Jamaican
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 (Non-American Hispanic)
36	"Spanish," non-European versus European unclear
37	Philippines/Guam
38	Spanish Filipino or Spanish Guamanian
39	Dominican/Mexican
40	Dominican/Puerto Rican
41	Dominican/Cent So Amer
42	Dominican/Cuban
43	Mexican/European Spanish
44	Puerto Rican/European Spanish
45	Cuban/European Spanish
46	Central or South American/European Spanish
47	Dominican/European Spanish
50	(All) Hispanic, white, no other information
51	(All) Hispanic, black, no other information
52	(All) Hispanic, Amer Indian, no other info
53	(All) Hispanic, Asian, no other information
54	(All) Hispanic, no other information
55	(All) Hispanic, Mezclada, Mezclado
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
68	Part Hispanic, part Asian, part white
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"
94 "Unknown"/"Don't Know"
97 "American" or "All of Them"

Appendix E: Model Summaries

Appendix E: Model Summaries

E.1 Introduction

The exhibits in this appendix list the covariates used in all the imputation models that were ran in the 2002 National Survey on Drug Use and Health (NSDUH)[139]. For each variable or set of variables to which the predictive mean neighborhood (PMN) imputation method was applied, two models were ran: 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; therefore, most of the exhibits are for only one age group.

The demographic variables are covered in Section E.2; Section E.3 deals with the drug variables. In this section, with the exception of the lifetime usage models, separate tables are provided for each drug-age group combination. Tables that cover the models for the household composition variables, derived from the questionnaire roster, are given in Section E.4. Section E.5 deals with the income variables, and Section E.6 provides tables for the health insurance models. Both of the methods that were used to create the final imputation-revised health insurance variables, the "Old Method" and the "Constituent Variables Method," are given in this section (see Chapter 10 for details).

In the exhibits, when the variables "age^2" and "age^3" are given, the superscripts represent squared and cubed, respectively. In these specific cases, the superscripts do not refer to footnotes. The variable "age" is the mean-centered age, where the age was "centered" by subtracting the mean age and where the mean was calculated across all respondents within the age group who were used to build the given model. The variables "age^2" and "age^3" represent the square and cube, respectively, of this mean-centered age variable. Also in the exhibits, when an asterisk "*" is given, it represents an interaction between two variables and not multiplication. In addition, when the abbreviation "MSA" is used, it represents "metropolitan statistical area."

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

E.2 Demographic Variables

Exhibit E.1 Summaries for Response Propensity Models (All Three Age Groups in Same Model)

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

Exhibit E.2 Summaries for Predictive Mean Models

Imputation Step	Variables Included in Predictive Mean Model
Race 12-17[1]	Census Region; Household Type; Age; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Race 18-25	Census Region; Household Type; Age; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Race 26+[2]	Census Region; Household Type; Age; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Hispanic Origin 12-17	Census Region; Imputation Revised Race; Household Type; Age; Age^2; Age^3; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Hispanic Origin 18-25	Census Region; Imputation Revised Race; Household Type; Age; Age^2; Age^3; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Hispanic Origin 26+	Census Region; Imputation Revised Race; Household Type; Age; Age^2; Age^3; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Marital Status[3]	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Population Density; Age; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Age*Gender
Hispanic Group[4]	Census Region; Imputation Revised Race; Gender; Age; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Age*Gender
Education Level 12-17[5]	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Education Level 18+	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Age; Age^2; Age^3; Age*Gender; Age^2*Gender; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation Revised Marital Status
Employment Status 15-25[6]	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Age; Age^2; Age*Gender; Age^2*Gender; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Employment Status 26+	Census Region; Imputation Revised Race; Imputation Revised Hispanic Origin Indicator; Gender; Age; Age*Gender; Percent Non-Hispanic Black in Segment; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation Revised Marital Status

[1]In the race predictive mean model for the 12-17 age group, household type was collapsed into a two-level covariate to avoid the "Data Warning" message in SUDAAN® (SUDAAN® is a registered trademark of RTI International). See Section 4.4.2.3.2 for details.
[2]In the race predictive mean model for the 26+ age group, household type, percent Hispanic population, and percent non-Hispanic black population were collapsed into two-level covariates to avoid the "Data Warning" message in SUDAAN®. See Section 4.4.2.3.2 for details.
[3]All age groups were modeled together for the marital status predictive mean model. This was done so that more covariates could have been included in the models. See Section 4.4.6 for details.
[4]All age groups were modeled together for the Hispanic-origin group predictive mean model, so that more covariates could have been included in the models.
[5]The predictive mean model for education level had five levels for the 12-17 age group, but four levels for the other two age groups. See Section 4.4.7.2.2 for details.
[6]The predictive mean model for employment status for the 15-17 age group also included the 18-25 year old respondents. This was necessary since the number of 15-17 year olds was too small to create a sufficiently predictive model.

E.3 Drug Variables

Exhibit E.3 Lifetime Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; MSA; Census Region; Cigarette Lifetime Indicator
18 to 25	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; MSA; Census Region; Cigarette Lifetime Indicator
26+	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; MSA; Census Region; Cigarette Lifetime Indicator

Exhibit E.4 Cigarettes: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	Not applicable (N/A)	N/A
Recency	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A
30-day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Age; Age^2; Age^3; Age*Gender; Age*Race
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Cigarettes 30-Day Frequency
Age at First Daily Use	Race; Gender; Census Region; MSA; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Age at First Use for Cigarettes; Cigarettes 30-Day Frequency

Exhibit E.5 Cigarettes: 18 to 25 Year Olds

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

Exhibit E.6 Cigarettes: 26+ Year Olds

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

Exhibit E.7 Cigars: 12 to 17 Year Olds

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

Exhibit E.8 Cigars: 18 to 25 Year Olds

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

Exhibit E.9 Cigars: 26+ Year Olds

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

Exhibit E.10 Pipes: 12 to 17 Year Olds

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

Exhibit E.11 Pipes: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, and Cigars; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Race; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Alcohol, Inhalants, and Marijuana	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, 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 E.12 Pipes: 26+ Year Olds

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

Exhibit E.13 Smokeless Tobacco (Chewing Tobacco and Snuff): 12 to 17 Year Olds

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

(continued)

Exhibit E.13 Smokeless Tobacco (Chewing Tobacco and Snuff): 12 to 17 Year Olds (continued)

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

Exhibit E.14 Smokeless Tobacco (Chewing Tobacco and Snuff): 18 to 25 Year Olds

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

(continued)

Exhibit E.14 Smokeless Tobacco (Chewing Tobacco and Snuff): 18 to 25 Year Olds (continued)

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

Exhibit E.15 Smokeless Tobacco (Chewing Tobacco and Snuff): 26+ Year Olds

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

(continued)

Exhibit E.15 Smokeless Tobacco (Chewing Tobacco and Snuff): 26+ Year Olds (continued)

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

Exhibit E.16 Alcohol: 12 to 17 Year Olds

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

Exhibit E.17 Alcohol: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, and Pipes; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Alcohol Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Alcohol Indicator
30-day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Alcohol 12-Month Frequency	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Alcohol 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.17 Alcohol: 18 to 25 Year Olds (continued)

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

Exhibit E.18 Alcohol: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, and Pipes; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Age Category; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Alcohol Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Alcohol Indicator
30-day Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, and Pipes; Imputation-Revised Lifetime Indicators for Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Alcohol 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.18 Alcohol: 26+ Year Olds (continued)

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

Exhibit E.19 Inhalants: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalants Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalants Indicator
30-day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Inhalants 12-Month Frequency	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Inhalants 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.19 Inhalants: 12 to 17 Year Olds (continued)

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

Exhibit E.20 Inhalants: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Gender; Census Region; MSA; Imputation-Revised Recencies for Cigars and Alcohol; Imputation-Revised Lifetime Indicators for Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalants Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalants Indicator
30-day Frequency	Imputation-Revised Recencies for Smokeless Tobacco and Pipes; Imputation-Revised Lifetime Indicators for Sedatives, Crack, and Heroin	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Inhalants 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.20 Inhalants: 18 to 25 Year Olds (continued)

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

Exhibit E.21 Inhalants: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Age Category; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Race; Imputation-Revised Recency for Alcohol; Imputation-Revised Lifetime Indicators for Cigars, Marijuana, Pain Relievers, Tranquilizers, Stimulants, Cocaine, and Crack	Age; Education Level; Imputation-Revised Lifetime Indicators for Pipes, Hallucinogens, Tranquilizers, Stimulants, Sedatives, and Crack
12-Month Frequency	Race; MSA; Imputation-Revised Recency for Alcohol; Imputation-Revised Lifetime Indicator for Marijuana	Age; Age^2; Age^3; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Cigars, and Alcohol; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Inhalants Indicator
30-day Frequency	Race; MSA	Marital Status; Census Region; Imputation-Revised Recency for Alcohol; Imputation-Revised Lifetime Indicators for Hallucinogens and Pain Relievers
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, and Alcohol; Inhalants 12-Month Frequency; Inhalants 30-Day Frequency

Exhibit E.22 Marijuana: 12 to 17 Year Olds

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

Exhibit E.23 Marijuana: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Marijuana Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Marijuana Indicator
30-day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Marijuana 12-Month Frequency	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Marijuana 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.23 Marijuana: 18 to 25 Year Olds (continued)

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

Exhibit E.24 Marijuana: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Age Category; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Marijuana Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Marijuana Indicator
30-day Frequency	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, and Inhalants; Imputation-Revised Lifetime Indicators for Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Marijuana 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.24 Marijuana: 26+ Year Olds (continued)

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

Exhibit E.25 Hallucinogens: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Hallucinogens Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Hallucinogens Indicator

(continued)

Exhibit E.25 Hallucinogens: 12 to 17 Year Olds (continued)

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

Exhibit E.26 Hallucinogens: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Marital Status; Education Level; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Hallucinogens Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Hallucinogens Indicator

(continued)

Exhibit E.26 Hallucinogens: 18 to 25 Year Olds (continued)

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

Exhibit E.27 Hallucinogens: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, and Marijuana; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Age Category; Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Race; Marital Status; Education Level; Employment Status; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Gender; Marital Status; Education Level; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Census Region; MSA; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack	Education Level; Imputation-Revised Recencies for Pipes and Inhalants; Imputation-Revised Lifetime Indicator for Heroin
30-day Frequency	Race; Gender; Imputation-Revised Recencies for Cigarettes and Marijuana; Imputation-Revised Lifetime Indicators for Stimulants and Cocaine	Gender; Race; Marital Status; Education Level; Employment Status; Imputation-Revised Lifetime Indicators for Pain Relievers, Tranquilizers, Cocaine, and Heroin

(continued)

Exhibit E.27 Hallucinogens: 26+ Year Olds (continued)

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

Exhibit E.28 Pain Relievers: 12 to 17 Year Olds

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

Exhibit E.29 Pain Relievers: 18 to 25 Year Olds

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

Exhibit E.30 Pain Relievers: 26+ Year Olds

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

Exhibit E.31 Tranquilizers: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Tranquilizers Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Tranquilizers Indicator
30-day Frequency	N/A	N/A

(continued)

Exhibit E.31 Tranquilizers: 12 to 17 Year Olds (continued)

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

Exhibit E.32 Tranquilizers: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Tranquilizers Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Tranquilizers Indicator
30-day Frequency	N/A	N/A

(continued)

Exhibit E.32 Tranquilizers: 18 to 25 Year Olds (continued)

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

Exhibit E.33 Tranquilizers: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Age Category; Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pain Relievers, and Crack	Age; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Gender; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, and Pipes; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, and Crack; Intermediate Past Month Tranquilizers Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, and Pain Relievers; Imputation-Revised Lifetime Indicators for Stimulants, Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Tranquilizers Indicator
30-day Frequency	N/A	N/A

(continued)

Exhibit E.33 Tranquilizers: 26+ Year Olds (continued)

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

Exhibit E.34 Stimulants: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Stimulants Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Stimulants Indicator
30-day Frequency	N/A	N/A

(continued)

Exhibit E.34 Stimulants: 12 to 17 Year Olds (continued)

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

Exhibit E.35 Stimulants: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Stimulants Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Intermediate Past Month Stimulants Indicator
30-day Frequency	N/A	N/A

(continued)

Exhibit E.35 Stimulants: 18 to 25 Year Olds (continued)

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

Exhibit E.36 Stimulants: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Age Category; Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Sedatives, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Age Category; Marital Status; Education Level; Census Region; MSA; State Rank; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Inhalants, Tranquilizers, Sedatives, Crack, and Heroin	Employment Status; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, and Cocaine
12-Month Frequency	Age Category; Gender; MSA; Imputation-Revised Recencies for Cigars and Marijuana; Imputation-Revised Lifetime Indicators for Sedatives, Crack, and Heroin; Intermediate Past Month Stimulants Indicator	Age; Age^2; Gender; Age*Gender; Marital Status; Education Level; Census Region; Imputation-Revised Recencies for Cigars, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Sedatives and Cocaine
30-day Frequency	N/A	N/A
Age at First Use	Age Category; Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, and Tranquilizers; Stimulants 12-Month Frequency

Exhibit E.37 Sedatives: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Marijuana, and Hallucinogens; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin	Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
12-Month Frequency	Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, and Tranquilizers; Imputation-Revised Lifetime Indicator for Cocaine; Intermediate Past Month Sedatives Indicator	Age; Gender; Race; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Intermediate Past Month Sedatives Indicator
30-day Frequency	N/A	N/A
Age at First Use	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicator for Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Sedatives 12-Month Frequency

Exhibit E.38 Sedatives: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Marital Status; Education Level; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Pipes, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Marital Status; Education Level; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Pipes, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
12-Month Frequency	Race; Gender; MSA; Imputation-Revised Recencies for Hallucinogens and Tranquilizers; Imputation-Revised Lifetime Indicators for Cocaine and Crack; Intermediate Past Month Sedatives Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Imputation-Revised Lifetime Indicators for 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 Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Marijuana, Pain Relievers, and Stimulants; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Sedatives 12-Month Frequency

Exhibit E.39 Sedatives: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Race; Marital Status; Education Level; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes and Alcohol; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, and Heroin	Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
Recency: past month vs. past year not past month	Gender; Race; Gender*Race; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Cigars, Pain Relievers, and Tranquilizers	Age; Gender; Age*Gender; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Cocaine, Crack, and Heroin
12-Month Frequency	Age Category; Race; Gender; Imputation-Revised Recency for Alcohol	Gender; Employment Status; Census Region; Imputation-Revised Recencies for Cigarettes, Marijuana, Hallucinogens, and Stimulants; Imputation-Revised Lifetime Indicators for Crack and Heroin
30-day Frequency	N/A	N/A
Age at First Use	Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Sedatives; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Sedatives 12-Month Frequency

Exhibit E.40 Cocaine: 12 to 17 Year Olds

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

Exhibit E.41 Cocaine: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Intermediate Past Month Cocaine Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Intermediate Past Month Cocaine Indicator
30-day Frequency	Race; Gender; Census Region; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, and Stimulants; Imputation-Revised Lifetime Indicators for Crack and Heroin	Gender; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Intermediate Cocaine 12-Month Frequency; Age; Age^2; Age^3; Age*Gender; Age*Race

(continued)

Exhibit E.41 Cocaine: 18 to 25 Year Olds (continued)

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

Exhibit E.42 Cocaine: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Gender; Race; Age^2; Age^3; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Marital Status; Employment Status; Education Level; Gender*Race; Age*Gender; Age*Race; State Rank; MSA; Census Region
Recency	Gender; Age Category; Race; Gender*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin	Age; Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Crack, and Heroin
12-Month Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Intermediate Past Month Cocaine Indicator	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Lifetime Indicators for Crack and Heroin; Intermediate Past Month Cocaine Indicator
30-day Frequency	Race; Gender; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Cigars, Hallucinogens, Pain Relievers, and Tranquilizers; Imputation-Revised Lifetime Indicators for Crack and Heroin	Education Level; Employment Status; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Hallucinogens, and Pain Relievers

(continued)

Exhibit E.42 Cocaine: 26+ Year Olds (continued)

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

Exhibit E.43 Heroin: 12 to 17 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Intermediate Lifetime Indicator for Chewing Tobacco; State Rank; Census Region
Recency: past year vs. not past year	Race; MSA; State Rank; Imputation-Revised Recency for Alcohol; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Stimulants, and Crack	Race; Census Region; MSA; State Rank; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Pain Relievers, Tranquilizers, Stimulants, Sedatives, and Crack
Recency: past month vs. past year not past month	Race; MSA; Imputation-Revised Lifetime Indicators for Pipes, Inhalants, Pain Relievers, Stimulants, and Crack	Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Tranquilizers, and Sedatives
12-Month Frequency	Imputation-Revised Recencies for Smokeless Tobacco, Pipes, and Inhalants	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Cigars, Pipes, Alcohol, and Inhalants
30-day Frequency	Census Region	Gender; Race; MSA
Age at First Use	Race; Imputation-Revised Recencies for Smokeless Tobacco, Pipes, Inhalants, Tranquilizers, and Stimulants	Age; Race; Census Region; MSA; Imputation-Revised Recencies for Alcohol, Hallucinogens, and Stimulants; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Alcohol, and Stimulants

Exhibit E.44 Heroin: 18 to 25 Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Age; Race; Intermediate Lifetime Indicators for Chewing Tobacco, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, and Stimulants; Marital Status; Employment Status; Education Level; State Rank; MSA; Census Region
Recency: past year vs. not past year	Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recency for Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Pipes, Inhalants, Tranquilizers, Stimulants, Sedatives, and Crack	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; State Rank; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack
Recency: past month vs. past year not past month	Marital Status	Age; Age^2; Age^3; Gender; Gender*Race; Age*Gender; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Smokeless Tobacco, Cigars, Pipes, Tranquilizers, Stimulants, and Sedatives
12-Month Frequency	Gender; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Marijuana, Hallucinogens, Pain Relievers, and Crack	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Race; Education Level; State Rank; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Inhalants, Tranquilizers, Stimulants, and Cocaine
30-day Frequency	Race; Census Region; Imputation-Revised Recency for Cigarettes	Gender; Race; Marital Status; Education Level; Employment Status; Census Region; MSA
Age at First Use	Imputation-Revised Recencies for Cigarettes and Alcohol	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Heroin 12-Month Frequency; Heroin 30-Day Frequency

Exhibit E.45 Heroin: 26+ Year Olds

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	*See Exhibit E.2*	Gender; Race; Cigarette Lifetime Indicator; Intermediate Lifetime Indicators for Snuff, Chewing Tobacco, Cigars, Pipes, and Alcohol; Marital Status; Employment Status; Education Level; Gender*Race; State Rank; MSA; Census Region
Recency: past year vs. not past year	Race; Marital Status; Education Level; Employment Status; Imputation-Revised Recencies for Cigarettes, Alcohol, and Marijuana; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Inhalants, Hallucinogens, Pain Relievers, Cocaine, and Crack	Age; Gender; Marital Status; Education Level; Employment Status; Census Region; State Rank; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Inhalants, Hallucinogens, Tranquilizers, Stimulants, Sedatives, and Crack
Recency: past month vs. past year not past month	Race; Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Inhalants, Hallucinogens, Cocaine, and Crack	Imputation-Revised Recency for Cigarettes; Imputation-Revised Lifetime Indicators for Cigars, Pipes, Inhalants, Pain Relievers, and Sedatives
12-Month Frequency	Age Category; Race; Gender; Imputation-Revised Recencies for Cigarettes, Cigars, Alcohol, and Marijuana	Age; Gender; Race; Gender*Race; Age*Gender; Age*Race; Marital Status; Education Level; Employment Status
30-day Frequency	Race; Gender; Census Region; Imputation-Revised Recencies for Cigarettes, Cigars, and Marijuana	Race; Marital Status; Imputation-Revised Recency for Marijuana
Age at First Use	Gender; Imputation-Revised Recencies for Cigarettes, Alcohol, Marijuana, and Cocaine	Age; Gender; Race; State Rank; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Marital Status; Education Level; Employment Status; Census Region; MSA; Imputation-Revised Recencies for Cigarettes, Smokeless Tobacco, Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin; Imputation-Revised Ages at First Use for Cigarettes, Cigarette Daily, Smokeless Tobacco, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, and Crack; Heroin 12-Month Frequency; Heroin 30-Day Frequency

E.4 Household Composition Variables

Exhibit E.46 Household Composition: 12 to 17 Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Age; Gender; Race; Age*Gender; Age*Race; Census Region; MSA; Percent Hispanic in Segment	Age; Total People in Household (Screener); Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Number of Persons Greater Than 64 Years Old in Household (HH65)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Other family present in Household (FAMSKIP)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment	Race; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Exhibit E.47 Household Composition: 18 to 25 Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Age; Race	Age; Total People in Household (Screener); Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Number of Persons Greater Than 64 Years Old in Household (HH65)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other family present in Household (FAMSKIP)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Gender; Race; Gender*Race; Age*Gender; Census Region; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percent Hispanic in Segment

Exhibit E.48 Household Composition: 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status; Total People in Household (Screener)	Age; Total People in Household (Screener); Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; 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; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Number of Persons Greater Than 64 in Household (HH65)	Age; Age^2; Gender; Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other Family Present in Household (FAMSKIP)	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Education Level; Employment Status

Exhibit E.49 Household Composition: 65+ Year Olds

Variable Requiring Imputation	Variables Included in Response Propensity	Variables Included in Roster Model
Household Size (TOTPEOP)	Gender	Age; Total People in Household (Screener); Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Number of Persons Younger Than 18 Years Old in Household (KID17)	Number of Eligible 12 to 17 in Household (Screener)	Age; Number of Eligible 12 to 17 in Household (Screener); Imputation-Revised Household Size; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment
Number of Persons Greater Than 64 Years old in Household (HH65)	Age; Gender; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Census Region; Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; Employment Status
Other Family Present in Household (FAMSKIP)	Age; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Household Size; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Percent Owner Occupied in Segment; Marital Status; Education Level	Marital Status

E.5 Income Variables

Exhibit E.50 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12 to 17	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Age^3*Gender; Age^3*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
18 to 25	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*Race; Age^3*Gender; Age^3*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
26 to 64	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
65+	Gender; Race; 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

Exhibit E.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12 to 17 Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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; Income State Rank
Supplemental Security Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security
Wages	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support
Food Stamps	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income
Welfare Payments	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income

(continued)

Exhibit E.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12 to 17 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Services	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments
# Welfare Months	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps
Investment Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services
Child Support	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income

(continued)

Exhibit E.51 Dichotomous Income Indicators in Predictive Mean Modeling: 12 to 17 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Other Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages
Total Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps

Exhibit E.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18 to 25 Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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; Income State Rank; Marital Status; Education Level; Employment Status
Supplemental Security Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security
Wages	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support
Food Stamps	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income
Welfare Payments	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income

(continued)

Exhibit E.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18 to 25 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Services	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments
# Welfare Months	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps; Marital Status; Education Level; Employment Status
Investment Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services
Child Support	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income

(continued)

Exhibit E.52 Dichotomous Income Indicators in Predictive Mean Modeling: 18 to 25 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Other Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages
Total Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps; Marital Status; Education Level; Employment Status

Exhibit E.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security
Wages	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support
Food Stamps	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Greater Than 64 Years Old in Household; Income State Rank; Intermediate Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income
Welfare Payments	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Greater Than 64 Years Old in Household; Income State Rank; Intermediate Indicator whether Family Received Social Security
Welfare Services	Age; Gender; Race; Age^2; Age^3; Census Region; 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 Greater Than 64 Years Old in Household; Income State Rank; Marital Status; Education Level; Intermediate Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income

(continued)

Exhibit E.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
# Welfare Months	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps; Marital Status; Education Level; Employment Status
Investment Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services
Child Support	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income
Other Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages

(continued)

Exhibit E.53 Dichotomous Income Indicators in Predictive Mean Modeling: 26 to 64 Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps; Marital Status; Education Level; Employment Status

Exhibit E.54 Dichotomous Income Indicators in Predictive Mean Modeling: 65+ Year Olds

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Gender*Race; Age*Gender; Age*Race; Census Region; Percent Owner Occupied in Segment; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status
Supplemental Security Income	Age; Gender; Race; Gender*Race; Age*Gender; Age*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 Indicator whether Family Received Social Security
Wages	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Persons Greater Than 64 Years Old in Household; Income State Rank; Marital Status; Education Level; Employment Status; Intermediate Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support
Food Stamps	Gender; Census Region; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income
Welfare Payments	Census Region; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised Number of Adults in Household; Income State Rank; Education Level; Employment Status; Intermediate Indicator whether Family Received Supplemental Security Income
Welfare Services	Age; Gender; Race; Census Region; MSA; 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; Intermediate Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income
# Welfare Months	Age; MSA

(continued)

Exhibit E.54 Dichotomous Income Indicators in Predictive Mean Modeling: 65+ Year Olds (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Investment Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services
Child Support	Age; Gender; Race; Census Region; MSA; Percent Non-Hispanic Black in Segment; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Income State Rank; Marital Status; Education Level; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income
Other Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages
Total Income	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Intermediate Indicator whether Family Received Supplemental Security Income; Intermediate Indicator whether Family Received Welfare Payments; Intermediate Indicator whether Family Received Welfare Services; Intermediate Indicator whether Family Received Investment Income; Intermediate Indicator whether Family Received Child Support; Intermediate Indicator whether Family Received Wages; Intermediate Indicator whether Family Received Other Income; Intermediate Indicator whether Family Received Food Stamps; Marital Status; Education Level; Employment Status

Exhibit E.55 Income Finer Categories in Response Propensity Models

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
12 to 17	Gender; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; Age^2*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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous)
26 to 64	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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous)

(continued)

Exhibit E.55 Income Finer Categories in Response Propensity Models (continued)

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
65+	Gender; Race; Census Region; MSA; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Percent Owner Occupied in Segment; Marital Status; Education Level; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous)

Exhibit E.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; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status
26 to 64	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status

(continued)

Exhibit E.56 Income Finer Categories in Predictive Mean Models (continued)

Age Group	Variables Included in Income Models (Finer Categorization)
65+	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age*Race; Age^2*Gender; 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 Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Supplemental Security Income; Imputation-Revised Indicator whether Family Received Welfare Payments; Imputation-Revised Indicator whether Family Received Welfare Services; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Child Support; Imputation-Revised Indicator whether Family Received Wages; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Family Received Food Stamps; Imputation-Revised Family Income (Dichotomous); Marital Status; Education Level; Employment Status

E.6 Health Insurance Variables

Exhibit E.57 Health Insurance, Constituent Variables Method: Response Propensity Models

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
12 to 17	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (4 levels)
	Other Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (4 levels)
18 to 25	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Gender; Race
	Other Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; Marital Status; Education Level; Employment Status; MSA; Percent Owner Occupied in Segment; Income (4 levels)
26 to 64	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Gender; Race; Marital Status; Education Level; Employment Status; MSA; Percent Owner Occupied in Segment; Income (4 levels)
	Other Health Insurance[1]	Age; Gender; Race; Gender*Race; Age*Gender; Age*Race; Education Level; Employment Status; MSA; Percent Owner Occupied in Segment; Income (4 levels); Marital Status
65+	**Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance**	Age; Marital Status; MSA; Percent Owner Occupied in Segment; Income (4 levels)
	Other Health Insurance[1]	Age; Gender; Race; Gender*Race; Age*Gender; Age*Race; Education Level; Employment Status; MSA; Percent Owner Occupied in Segment; Income (2 levels); Marital Status

[1]The 26-64 and 65+ age groups were included in the same response propensity model for other health insurance.

Exhibit E.58 Health Insurance, Constituent Variables Method: Predictive Mean Models, 12 to 17 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Household Size
Medicare	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Imputation-Revised Indicator whether Family Received Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (3 levels); Imputation-Revised Indicator whether Family Received Other Income; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Household Size

Exhibit E.59 Health Insurance, Constituent Variables Method: Predictive Mean Models, 18 to 25 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size
Medicare[2]	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (3 levels); Personal Other Income; Lifetime Military Service; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size

Exhibit E.60 Health Insurance, Constituent Variables Method: Predictive Mean Models, 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Employment Status; Education Level; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size
Medicare[2]	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (3 levels); Personal Other Income; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Income (4 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance[3]	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size

Exhibit E.61 Health Insurance, Constituent Variables Method: Predictive Mean Models, 65+ Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age^2; Gender; Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Income (2 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Household Size
Medicare	Age; Age^2; Gender; Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Personal Social Security; Intermediate MEDICAID/CHIP Coverage
CHAMPUS	Age; Age^2; Gender; Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Income (2 levels); Personal Other Income; Lifetime Military Service; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage
Private Health Insurance	Age; Age^2; Gender; Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Income (2 levels); Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Family Received Other Income; Imputation-Revised Indicator whether Other Family Members in Household; Household Size; Intermediate MEDICAID/CHIP Coverage; Intermediate MEDICARE Coverage; Intermediate CHAMPUS Coverage
Other Health Insurance[3]	Age; Age^2; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Percent Owner Occupied in Segment; Imputation-Revised Indicator whether Family Received Wages; Indicator whether Family Participates in Government Assistance Programs; Imputation-Revised Indicator whether Family Received Social Security; Imputation-Revised Indicator whether Family Received Investment Income; Imputation-Revised Indicator whether Other Family Members in Household; Imputation-Revised Household Size

[2]The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.
[3]The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

Exhibit E.62 Old Method Health Insurance, Based on INSUR3: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
18 to 25	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
26 to 64	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
65+	Gender; Race; Percent Owner Occupied in Segment

Exhibit E.63 Old Method Health Insurance, Based on INSUR: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12 to 17	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
18 to 25	Age; Age^2; Age^3; Gender; Race; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
26 to 64	Gender; Race; Gender*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
65+	Gender; Race; Education Level; Percent Owner Occupied in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size

Exhibit E.64 Old Method Health Insurance: Predictive Mean Models, 12 to 17 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3

Exhibit E.65 Old Method Health Insurance: Predictive Mean Models, 18 to 25 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3

Exhibit E.66 Old Method Health Insurance: Predictive Mean Models, 26 to 64 Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3

Exhibit E.67 Old Method Health Insurance: Predictive Mean Models, 65+ Year Olds

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Overall Health Insurance (INSUR)	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Marital Status; Education Level; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size
Private Health Insurance[1]	Age; Gender; Race; Age^2; Age^3; Gender*Race; Age*Gender; Age^2*Gender; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Non-Hispanic Black in Segment; Imputation-Revised Household Size

[1]Item response definition based on INSUR3

Appendix F: Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

Appendix F: Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

F.1 Introduction

For all the 2002 National Survey on Drug Use and Health (NSDUH)[140] variables for which imputations were implemented using predictive mean neighborhoods (PMN), whether the method was univariate (UPMN) or multivariate (MPMN), restrictions were placed upon the neighborhood prior to the assignment of imputed values. The pool of potential donors for a given recipient was restricted so that donors and recipients were as alike as possible (likeness constraints), and the donor's values were consistent with the preexisting nonmissing values of the recipient (logical constraints). Logical constraints (summarized in Appendix G) were not loosened because this would have resulted in an inconsistency that would not have been countenanced.[141] However, some likeness constraints were loosened, even though this resulted in donors and recipients being less alike in various cases. If no donors were available under the most stringent set of constraints, the likeness constraints were loosened, one at a time, until a donor was found. This appendix summarizes the number of cases for which donors were available under each of the various likeness constraints, starting with the most stringent constraint. The appendix is organized by groups of variables requiring imputation using the PMN method: demographics, lifetime use of drugs, recency and frequency of drug use, age at first drug use, household roster, income, and health insurance. The labels for some of the likeness constraints given in the exhibits are not self-evident; therefore, more complete descriptions are given in the following paragraphs.

Although statistical imputation of the drug use or income variables could not have proceeded separately within each State due to insufficient pools of donors, information about the State of residence of each respondent was incorporated in the PMN procedure. For the drug use variables, in the hot-deck step of PMN, respondents were separated into three State usage-level categories for each drug depending on the response variable of interest. Respondents from States with high usage of a given drug were placed in one category, respondents from medium usage States into another, and the remainder into a third category. The States were separated into three income groups for the income variables, depending upon the proportion of families with incomes greater than or equal to $20,000. As with the drug use variables, respondents from high-income States (by this measure) were placed in one category, respondents from medium income states into another category, and the remainder into a third category. In the 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 of the

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

[141] Logical constraints define what is normally referred to as an "imputation class."

univariate imputation exhibits. In either case, this phrase represents one of the likeness constraints. It also defines the neighborhood. Once this constraint was loosened, the neighborhood was abandoned and the candidate with the predicted means closest to the recipient's, subject to the constraints that were still on the pool of donors, was chosen as the donor.

F.2 Demographics

F.2.1 Race Variables

Exhibit F.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	133	73	23
(A) Donor's predicted means within 5 percent of recipient's predicted means	413	448	321
None	73	151	119

F.2.2 Hispanic Origin Variables

Exhibit F.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	53	4	0
(A) Donor's predicted mean within 5 percent of recipient's predicted mean	57	4	5

F.2.3 Marital Status Variables

Exhibit F.3 Marital Status Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Donor’s age within 3 years of recipient’s age (B) Donor's predicted mean within 5 percent of recipient's predicted mean	4	3	6
(A) Donor’s age within 3 years of recipient’s age	0	1	0

F.2.4 Hispanic Group Variables

Exhibit F.4 Hispanic Group Imputations

Likeness Constraints	Frequency[1]
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	11
(A) Donor's predicted means within 5 percent of recipient's predicted means	63
None	10

[1]The hot-deck program for Hispanic Group was not separated into age groups.

F.2.5 Education Variables

Exhibit F.4 Education Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	0	1	0
(A) Donor's predicted means within 5 percent of recipient's predicted means	1	4	6
None	0	0	1

F.2.6 Employment Variables

Exhibit F.4 Employment Imputations

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Segment of donor = Segment of recipient (B) Donor's predicted means within 5 percent of recipient's predicted means	0	4	1
(A) Donor's predicted means within 5 percent of recipient's predicted means	5	6	31
None	0	0	5

F.3 Drug Variables

The imputation of the drug use variables was done separately for three age groups: 12 through 17, 18 through 25, and 26 or older. For each of the drugs, a multivariate imputation was done for the recency and frequency variables, and a univariate imputation was done for the age at first use variable(s). The exhibits in this appendix show the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

F.3.1 Likeness Constraints for Lifetime Imputation

Exhibit F.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	412	73	56
(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	60	48	41
(A) State rank of donor = State rank of recipient	36	20	26

F.3.2 Likeness Constraints for Recency and Frequency Imputation, by Drug

Exhibits F.6 through F.19 present information on the likeness constraints for recency and frequency imputation for the following drugs: tobacco (i.e., cigarettes, smokeless tobacco [chewing tobacco and snuff], cigars, and pipes), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, stimulants, and sedatives), cocaine, and heroin.

Exhibit F.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	472	123	11
(A) State rank of donor = State rank of recipient	26	9	10

Exhibit F.7 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	112	103	3
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	5	3	0
(A) State rank of donor = State rank of recipient	94	31	8

Exhibit F.8 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	275	195	47
(A) State rank of donor = State rank of recipient	41	17	10

Exhibit F.9 Pipe Recency Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted probability of past month use within 5 percent of recipient's predicted probability of past month use	1	2	1

Exhibit F.10 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	576	548	403
(A) State rank of donor = State rank of recipient	296	136	90

Exhibit F.11 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	70	3	0
(A) State rank of donor = State rank of recipient	217	56	12

Exhibit F.12 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	109	115	30
(A) State rank of donor = State rank of recipient	202	136	58

Exhibit F.13 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	18	170	18
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	36	69	16
(A) State rank of donor = State rank of recipient	208	214	43
None	1	0	0

Exhibit F.14 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	105	52	16
(A) State rank of donor = State rank of recipient	169	71	45

Exhibit F.15 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	8	16	1
(A) State rank of donor = State rank of recipient	41	48	19

Exhibit F.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	20	32	9
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	6	4	2
(A) State rank of donor = State rank of recipient	117	66	30
None	1	0	2

Exhibit F.17 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	3	1	0
(A) State rank of donor = State rank of recipient	22	9	6

Exhibit F.18 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	0	17	8
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	1	9	7
(A) State rank of donor = State rank of recipient	58	95	57
None	0	1*	0

*Due to problems with finding a donor who met the logical constraints, this item nonrespondent underwent a random imputation for cocaine 30-day frequency of use, before all other nonmissing cocaine recency and frequency variables were imputed.

Exhibit F.19 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	1	0
(A) State rank of donor = State rank of recipient	12	10	6
None	0	2	0

F.3.3 Likeness Constraints for Age at First Use Imputation, by Drug

Exhibits F.20 through F.33 present information on the likeness constraints for age at first use (AFU) imputation for the following drugs: tobacco (i.e., cigarettes, cigarette daily use, smokeless tobacco [chewing tobacco and snuff], and cigars), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, stimulants, and sedatives), cocaine, and heroin.

Exhibit F.20 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	408	138	108
(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	1	6
(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	2
(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 F.21 Cigarette Age at First Daily Use Imputation

Likeness Constraints	Frequency		
	12-17	18-25	26+
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient 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	31	31	74
(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	5	1	8
(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	0	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	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.22 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	125	142	53
(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	20	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 (these checks are only done for the drugs for which the recipient has missing AFU)	13	2	10
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient,* Age of donor ≥ Age of recipient	6	0	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.23 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	243	209	183
(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	16
(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	12
(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	1

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.24 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	373	170	211
(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	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	0	0	6
(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 F.25 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	271	80	31
(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	1	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	1	0	2
(A) Age of donor = Age of recipient	0	0	2
None	0	0	3

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.26 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	114	62	60
(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	4
(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	1

Exhibit F.27 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	66	75	43
(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	12	7	6
(A) Age of donor = Age of recipient (B) Donor's predicted mean within 5 percent of recipient's predicted mean	15	7	7
(A) Age of donor = Age of recipient	7	2	10
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens),* Age of donor ≥ Age of recipient	0	2	1
(A) AFU of donor ≤ Age of recipient (for overall hallucinogens)*	3	0	2

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.28 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	276	183	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	4	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	18
(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	4

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.29 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	47	53	46
(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	5	1	8
(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	3	6
(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 F.30 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	83	53	27
(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	10	4	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 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	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 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)	4	4	3
(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	0	3
(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)*	6	0	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.31 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	14	7	11
(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	5	3	4
(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	9	3	5
(A) Age of donor = Age of recipient	0	0	1

Exhibit F.32 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	21	26	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 (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	3	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 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	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 (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)	2	0	3
(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	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	0

* Although this is a logical constraint, it is included for the sake of clarity.

Exhibit F.33   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	6	2	2
(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	2
(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	1
(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

* Although this is a logical constraint, it is included for the sake of clarity.

F.4 Household Composition (Roster) Variables

Exhibits F.34 through F.37 present information on the likeness constraints applied during the imputation procedures for the four household composition (roster) variables, IRHHSIZE, IRKID17, IRHH65, and IRFAMSKP.

Exhibit F.34   IRHHSIZE Imputations

Likeness Constraints	Frequency			
	12-17	**18-25**	**26-64**	**65+**
Donor's predicted mean within 5 percent of recipient's predicted mean	9	10	38	3

Exhibit F.35 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	51	65	71	3
(A) IRHHSIZE of donor = IRHHSIZE of recipient	0	1	0	0
None	1*	1	0	0

* Due to problems with finding a donor, this item nonrespondent underwent a random imputation.

Exhibit F.36 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	162	97	70	3
(A) IRHHSIZE of donor = IRHHSIZE of recipient	3	2	3	0
None	1	1	0	0

Exhibit F.37 IRFAMSKP Imputations

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) IRKID17 of donor = IRKID17 of recipient	9	12	37	1
(A) IRKID17 of donor = IRKID17 of recipient	0	0	0	0

F.5 Income Variables

F.5.1 Binary Variable Phase

Six of the binary income variables were directly related to a respondent's socioeconomic status. Hence, if a recipient required imputation for one or more of these six variables (i.e., welfare payments, welfare services, food stamps, binary income, investment income, and months on welfare), but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables. In the tables, these six variables are referred to as "welfare-correlated variables." All of the other likeness constraints that were applied are self-explanatory in the tables.

Exhibit F.38 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 (E) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing)	1368	1239	441	69
(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) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing)	544	493	376	133
(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 (D) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing)	7	23	21	7

(continued)

Exhibit F.38 Binary Income Imputations (continued)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(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 (C) If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing)	0	0	6	1
(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 (B) If recipient is missing only one edited variable of a (personal, other-If recipient is missing months-on-welfare, then the donor must match the recipient with respect to personal welfare payments (if nonmissing), other-family welfare payments (if nonmissing), personal welfare services (if nonmissing), and other-family welfare services (if nonmissing)	11	18	12	4
(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 (B) If recipient is missing only one edited variable of a (personal, other-If recipient is missing months-on-welfare, then the donor must match the recipient with respect to family welfare payments (if nonmissing) and family welfare services (if nonmissing)	0	0	0	1

F.5.2 Specific Category Phase

Exhibit F.39 Specific Income 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) Edited specific category personal income of donor = Edited specific category personal income of recipient, if nonmissing (C) Edited specific category family income of donor = Edited specific category family income of recipient, if nonmissing	2912	2782	1637	409
(A) Donor's predicted mean within 5 percent of recipient's predicted mean (B) Edited specific category family income of donor ≥ Edited specific category personal income of recipient, if not missing* (C) Edited specific category personal income of donor ≤ Edited specific category family income of recipient, if not missing*	10	3	1	7

* Although this is a logical constraint, it is included for the sake of clarity.

F.6 Health Insurance Variables

Exhibit F.40 presents information on the likeness constraints for the health insurance variables created using the "Old Method." The remaining tables present information for the health insurance variables created using the "Constituent Variables Method." See Chapter 10 for an explanation of the two methods. Briefly, in the Constituent Variables Method, four variables (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT) were imputed simultaneously in an MPMN program, and one variable (IROTHHLT) was imputed in a UPMN program following the imputation of other four variables. For the MPMN, the likeness constraints, which were applied to the variables, differed between missingness patterns, and sometimes the constraints differed between age groups within the same missingness pattern. As a result, there is at least one table for each missingness pattern. The final table in this section (Exhibit F.64) presents the likeness constraints applied in the UPMN program for IROTHHLT.

In several instances in these health insurance tables, variable names are used without description for the purposes of brevity. (See Chapter 10 for greater details.) For the health insurance imputations, matches between donors and recipients were attempted on the nonmissing values of the variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. These variables are the edited indicators of whether the respondent received health insurance from Medicaid/State health insurance programs for children, Medicare, Champus, or private health insurance, respectively. These were the base variables used in the creation of the imputation-revised variables (IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT). In addition to the edited health insurance variables, other variables, which were used as likeness constraints, are only identified in the tables by their variable names. These include SERVICE (an indicator of whether the respondent had ever been in the military service), GOVTPROG (an indicator of whether the respondent's family participated in government public assistance programs), INCOME (a 4-level categorical family income variable, with levels <$20K, $20K-<$50K, $50K-<$75K, and $75K or over), IRFAMIN1 (a 2-level family income variable with levels <$20K and $20K or over), IRFAMOTH/IRPOTH (an indicator of whether the respondent's family in the household or the respondent himself/herself received income from sources other than those considered in the income questions of the questionnaire), and IRFAMSOC/IRPSOC (an indicator of whether the respondent's family in the household or the respondent himself/herself received income from social security). For the latter two sets of variables, the match between donors and recipient was attempted on the personal income variable if the respondent was 18 or older. However, if the respondent was under 18, the match was attempted on the family income variable.

Exhibit F.40 Health Insurance (IRINSUR, IRINSUR3) and Private Health Insurance (IRPINSUR) Imputations, Old Method

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Age of donor = Age of recipient (B) Donor's predicted means each within 5 percent of recipient's predicted means	396	201	43	10
(A) Age of donor = Age of recipient	16	1	5	1
None	0	0	0	0

Exhibit F.41 Health Insurance, Constituent Variables Method (MPMN), Only CAIDCHIP Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for CAIDCHIP within 5 percent of recipient's predicted mean (B) GOVTPROG of donor = GOVTPROG of recipient (C) MEDICARE, CHAMPUS, and PRVHLTIN of donor = MEDICARE, CHAMPUS, and PRVHLTIN of recipient	129	72	13	8
(A) Donor's predicted mean for CAIDCHIP within 5 percent of recipient's predicted mean (B) GOVTPROG of donor = GOVTPROG of recipient	1	0	1	0
(A) Donor's predicted mean for CAIDCHIP within 5 percent of recipient's predicted mean	0	0	0	0

Exhibit F.42 Health Insurance, Constituent Variables Method (MPMN), Only MEDICARE Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for MEDICARE within 5 percent of recipient's predicted mean (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) CAIDCHIP, CHAMPUS, and PRVHLTIN of donor = CAIDCHIP, CHAMPUS, and PRVHLTIN of recipient	13	23	7	1
(A) Donor's predicted mean for MEDICARE within 5 percent of recipient's predicted mean (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient	2	0	0	0
(A) Donor's predicted mean for MEDICARE within 5 percent of recipient's predicted mean (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	0	0	0	0

Exhibit F.43 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and MEDICARE Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) CHAMPUS and PRVHLTIN of donor = CHAMPUS and PRVHLTIN of recipient	8	12	0	0
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient	1	1	1	1
(A) Donor's predicted means for CAIDCHIP and MEDICARE within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	2	2	0	0

Exhibit F.44 Health Insurance, Constituent Variables Method (MPMN), Only CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean for CHAMPUS within 5 percent of recipient's predicted mean (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) CAIDCHIP, MEDICARE, and PRVHLTIN of donor = CAIDCHIP, MEDICARE, and PRVHLTIN of recipient	21	20	6	2
(A) Donor's predicted mean for CHAMPUS within 5 percent of recipient's predicted mean (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	2	0	0	0
(A) Donor's predicted mean for CHAMPUS within 5 percent of recipient's predicted mean (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0	0

Exhibit F.45 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) MEDICARE and PRVHLTIN of donor = MEDICARE and PRVHLTIN of recipient	22	9	1	0
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	4	2	0	0
(A) Donor's predicted means for CAIDCHIP and CHAMPUS within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	0	0	0

Exhibit F.46 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for MEDICARE and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (F) CAIDCHIP and PRVHLTIN of donor = CAIDCHIP and PRVHLTIN of recipient	2	0	0	0
(A) Donor's predicted means for MEDICARE and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	1	0	0	0
(A) Donor's predicted means for MEDICARE and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	1	0	0

Exhibit F.47 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and CHAMPUS Missing

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (G) PRVHLTIN of donor = PRVHLTIN of recipient	5	1	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient	2	2	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and CHAMPUS within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	6	2	1	0

Exhibit F.48 Health Insurance, Constituent Variables Method (MPMN), Only PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) INCOME of donor = INCOME of recipient (C) CAIDCHIP, MEDICARE, and CHAMPUS of donor = CAIDCHIP, MEDICARE, and CHAMPUS of recipient	112	89	6
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) INCOME of donor = INCOME of recipient	4	1	0
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) IRFAMIN1 of donor = IRFAMIN1 of recipient (this constraint does not apply to 65+)	0	0	0
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean	0	0	0

Exhibit F.49 Health Insurance, Constituent Variables Method (MPMN), Only PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) IRFAMIN1 of donor = IRFAMIN1 of recipient (C) CAIDCHIP, MEDICARE, and CHAMPUS of donor = CAIDCHIP, MEDICARE, and CHAMPUS of recipient	4
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean (B) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted mean for PRVHLTIN within 5 percent of recipient's predicted mean	0

Exhibit F.50 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) INCOME of donor = INCOME of recipient (D) MEDICARE and CHAMPUS of donor = MEDICARE and CHAMPUS of recipient	50	9	1
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) INCOME of donor = INCOME of recipient	3	1	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) IRFAMIN1 of donor = IRFAMIN1 of recipient	2	0	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means	0	0	0
None	5	0	0

Exhibit F.51 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient (D) MEDICARE and CHAMPUS of donor = MEDICARE and CHAMPUS of recipient	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CAIDCHIP and PRVHLTIN within 5 percent of recipient's predicted means	0

Exhibit F.52 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26-64
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) INCOME of donor = INCOME of recipient (E) CAIDCHIP and CHAMPUS of donor = CAIDCHIP and CHAMPUS of recipient	1	0	0
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) INCOME of donor = INCOME of recipient	0	1	0
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	0	0
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	0	0	0

(continued)

Exhibit F.52 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Youngest Three Age Groups (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	1	0	0

Exhibit F.53 Health Insurance, Constituent Variables Method (MPMN), MEDICARE and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient (D) CAIDCHIP and CHAMPUS of donor = CAIDCHIP and CHAMPUS of recipient	0
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means (B) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for MEDICARE and PRVHLTIN within 5 percent of recipient's predicted means	0

Exhibit F.54 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) INCOME of donor = INCOME of recipient (F) CHAMPUS of donor = CHAMPUS of recipient	5	2	0

(continued)

Exhibit F.54 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and PRVHLTIN Missing, Youngest Three Age Groups (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	1	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	0	1	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14	6	2	1

Exhibit F.55 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRPSOC of donor = IRPSOC of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient (E) CHAMPUS of donor = CHAMPUS of recipient	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRPSOC of donor = IRPSOC of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, and PRVHLTIN within 5 percent of recipient's predicted means	0

Exhibit F.56 Health Insurance, Constituent Variables Method (MPMN), CHAMPUS and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) INCOME of donor = INCOME of recipient (E) CAIDCHIP and MEDICARE of donor = CAIDCHIP and MEDICARE of recipient	7	5	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (D) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	1	2	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	0	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	2	0	0

Exhibit F.57 Health Insurance, Constituent Variables Method (MPMN), CHAMPUS and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRPOTH of donor = IRPOTH of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient (E) CAIDCHIP and MEDICARE of donor = CAIDCHIP and MEDICARE of recipient	1
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRPOTH of donor = IRPOTH of recipient (D) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CHAMPUS and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	1

Exhibit F.58 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) INCOME of donor = INCOME of recipient (F) MEDICARE of donor = MEDICARE of recipient	32	2	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (E) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing) (C) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	0	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	1	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	25	10	0

Exhibit F.59 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient (F) MEDICARE of donor = MEDICARE of recipient	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for CAIDCHIP, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	0

Exhibit F.60 Health Insurance, Constituent Variables Method (MPMN), MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (F) INCOME of donor = INCOME of recipient (G) CAIDCHIP of donor = CAIDCHIP of recipient	1	0	0

(continued)

Exhibit F.60 Health Insurance, Constituent Variables Method (MPMN), MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (F) INCOME of donor = INCOME of recipient	0	0	0
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMIN1 of donor = IRFAMIN1 of recipient	0	0	0
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	1	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	2	1	0

Exhibit F.61 Health Insurance, Constituent Variables Method (MPMN), MEDICARE, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency 65+
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) IRPSOC of donor = IRPSOC of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient (F) CAIDCHIP of donor = CAIDCHIP of recipient	0
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) IRPSOC of donor = IRPSOC of recipient (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRPOTH of donor = IRPOTH of recipient (E) IRFAMIN1 of donor = IRFAMIN1 of recipient	0
(A) Donor's predicted means for MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	0

Exhibit F.62 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) GOVTPROG of donor = GOVTPROG of recipient (D) IRFAMSOC/IRPSOC of donor = IRFAMSOC/IRPSOC of recipient (E) SERVICE of donor = SERVICE of recipient (if nonmissing) (F) IRFAMOTH/IRPOTH of donor = IRFAMOTH/IRPOTH of recipient (G) INCOME of donor = INCOME of recipient	37	4	5

(continued)

Exhibit F.62 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN Missing, Youngest Three Age Groups (continued)

Likeness Constraints	Frequency		
	12-17	18-25	26-64
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing) (D) IRFAMIN1 of donor = IRFAMIN1 of recipient	1	1	0
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (C) SERVICE of donor = SERVICE of recipient (if nonmissing)	0	0	0
(A) If the recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then: a. If the recipient has no job due to disability (JBSTATR=14), then the donor must also have no job due to disability b. If the recipient has JBSTATR not equal to 14, then the donor must also have JBSTATR not equal to 14 (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	42	12	28

Exhibit F.63 Health Insurance, Constituent Variables Method (MPMN), CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN Missing, Oldest Age Group

Likeness Constraints	Frequency
	65+
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) GOVTPROG of donor = GOVTPROG of recipient (C) IRPSOC of donor = IRPSOC of recipient (D) SERVICE of donor = SERVICE of recipient (if nonmissing) (E) IRPOTH of donor = IRPOTH of recipient (F) IRFAMIN1 of donor = IRFAMIN1 of recipient	1
(A) Donor's predicted means for CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN within 5 percent of recipient's predicted means (B) SERVICE of donor = SERVICE of recipient (if nonmissing)	0
(A) SERVICE of donor = SERVICE of recipient (if nonmissing)	3

Exhibit F.64 Health Insurance, Constituent Variables Method (UPMN), Any Other Health Insurance

Likeness Constraints	Frequency		
	12-17	18-25	26+
Donor's predicted mean within 5 percent of recipient's predicted mean	62	89	14
None	3	2	1

Appendix G: Missingness Patterns

Appendix G: Missingness Patterns

G.1 Introduction

For the majority of variables that had missing values imputed in the 2002 National Survey on Drug Use and Health (NSDUH),[142] the imputation method used was Predictive Mean Neighborhoods (PMN). Some of these variables were imputed in sets. Specifically, an item nonrespondent with missing values for more than one variable in the set received values for all missing variables from the same donor. This is referred to as a "multivariate assignment." On the other hand, some variables were imputed one at a time using a "univariate assignment." In addition, some of the variables were imputed using a predictive mean vector with more than one element (multivariate matching), while others were imputed using a predictive mean vector with only one element (univariate matching). For variables that were binary or continuous and were not part of a multivariate set, the predictive mean vector and the assignment of imputed values were both univariate. However, multinomial variables that were not part of a multivariate set would have been imputed using a multivariate vector of predicted means (from a multinomial logistic model), from which a single imputed value (the level of the categorical variable) would have been imputed. A multivariate set of variables could have been imputed based on a single univariate model. This could have occurred if the variables were all inextricably related, whereby a model from one of the variables would have been sufficient to describe the responses for all the characteristics of interest. In most cases, a multivariate predictive mean vector was used to match donors and recipients for a multivariate set of response variables. Exhibit G.1 provides examples of variables that were imputed using each of the four methods.

Exhibit G.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, IRMJAGE	{IRPINC2, IRFINC2, IRFAMIN2}, {IRCOCAGE, IRCRKAGE}
Predictive mean vector has more than one element (multivariate matching)	IRMARIT, IRHOGRP3, EMPSTATY, IREDUC	{IRRACE, IRNWRACE}, {lifetime drug use}, {IRHERRC, IRHERFY, IRHERFM}, {binary sources of income}, {IRINSUR, IRINSUR3, IRPINSUR}, {IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT}

For many of these variables, the item nonrespondents were segregated into missingness patterns, which were simply patterns of nonresponse. Missingness patterns arose in two ways.

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

The first occurred for sets of variables that underwent multivariate assignment: item nonrespondents were segregated into missingness patterns based on which variables were missing. The second way occurred when logical editing restricted an item nonrespondent to only a subset of the variable's possible values. For example, logical editing sometimes restricted a lifetime user of a drug to past year use; in these cases, the recipient received a final imputed value of 1 or 2 for drug recency. This could have happened for any variable(s) that underwent multivariate matching.

This appendix focuses on the variables, or sets of variables, for which the set of logical constraints and/or the predictive mean vector differed between missingness patterns. It is limited to variables to which the PMN method was applied. The other imputation methods used in the 2002 survey were not multivariate. The exhibits in this appendix specify, for each missingness pattern:

1) the number of item nonrespondents exhibiting the pattern ("Number of Cases");

2) the set of logical constraints applied to the potential donors ("Logical Constraints");

3) the elements of the predictive mean vector ("Predictive Mean Vector") used to calculate the Mahalanobis distance from recipient to potential donor, as well as to restrict the donor set via the delta constraints as described in Appendix F.

Often, differences between missingness patterns with respect to the predictive mean vector were due to the use of conditional probabilities. If something about the item nonrespondent was known, probabilities, conditioned on what was known, were used. For example, only past month users were included in models for 30-day frequency. Therefore, the predictive means calculated using these models were conditional on past month use of the drug. If an item nonrespondent was missing both recency and 30-day frequency for that drug, probabilities conditional on lifetime use, not on past month use, were used for the predictive mean vector. Conditional probabilities often resulted if the variables, which were imputed using a multivariate assignment method, were related in a hierarchical manner, such as overall health insurance and private health insurance in the "Old Method" (see Chapter 10 for details). Also, these types of conditional probabilities occurred if partial information was available about an item nonrespondent, such as the cases where it was known that the recipient was a past year user of a drug, but it was unknown whether he or she was a past month user.

Section G.2 shows the variable or set of variables that used missingness patterns along with logical constraints and predictive mean vectors, as appropriate. Some exhibits also give the number of item nonrespondents showing each missingness pattern. Section G.2.1 deals with employment status, Section G.2.2 deals with drug lifetime use, Section G.2.3 focuses on drug recency and frequency, Section G.2.4 is concerned with the source of income variables, and Sections G.2.5 presents information on the health insurance variables.

G.2 Exhibits Showing Missingness Patterns and the Restrictions on the Set of Potential Donors

A few items to note regarding the exhibits in Section G.2 are as follows. In the missingness pattern section, no entry in the columns indicates that all information was available; an entry of "Missing" indicates that all information was missing. Other entries in the missingness pattern section give the available information, indicating that the information was partially missing. However, if the entry is given in parentheses, all information was present and additional details are given in the respective exhibit.

G.2.1 Employment Status

Conditional probabilities were used for employment status for the first time in the 2002 NSDUH. Exhibit G.2 illustrates the two missingness patterns for employment status.

Exhibit G.2 Restrictions and Portion of the Predictive Mean Vector for Employment Status

#	Employment Status	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Completely missing	36	None	1. E1 2. E2 3. E3
2	Known to be employed; part-time vs. full-time status unknown	16	Donor must be employed	1. E1/(E1+E2)

[1]The predictive mean vector components are defined by the following:
1. E1 = P(employed full time)
2. E2 = P(employed part time)
3. E3 = P(unemployed)

G.2.2 Drug Lifetime Use

There were a large number of missingness patterns for drug lifetime use. The response to the gate question for cigarettes must have been nonmissing for the survey to have been considered complete, but any combination of the other lifetime drug variables may have been missing. There were 14 other gate questions in the 2002 questionnaire, plus several subgate questions.

There were no logical constraints for any of these missingness patterns.

The probabilities associated with the 14 gate questions (Exhibit G.3) formed the full predictive mean vector. Only the probabilities associated with the gate questions, for which the responses were missing, were used in the predictive mean vector for each item nonrespondent.

Exhibit G.3 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)

G.2.3 Drug Recency and Frequency

Exhibits G.4 to G.21 on the following pages illustrate missingness patterns for drug recency and frequency of use. In this section, pain relievers, sedatives, and tranquilizers had identical missingness patterns and are therefore presented in the same exhibit. Many exhibits in this section abbreviate certain words. "Recency" is an abbreviation for "Recency of Use," "Frequency" or "Freq" is an abbreviation for "Frequency of Use," and "30-day binge drink" or "DR5DAY" is an abbreviation for the "number of days in the past 30 days when the respondent consumed five or more alcoholic drinks."

Exhibit G.4 Constraints for Tobacco (Cigarettes and Cigars)

Constraint #	Logical Constraint
Tob1	If the difference between the recipient's current age and his or 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 or her current age, the donor's 30-day frequency (1) cannot be greater than the number of days between the recipient's interview date and his or her date of first drug use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his or her birthday (inclusive)

Exhibit G.5   Restrictions and Portion of the Predictive Mean Vector for Cigarette Users

Missingness Pattern			Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Recency	30-Day Frequency			
1	Past year	Missing	17	(Tob1), (Tob5)	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)
2	Missing (lifetime use known)	Missing	28	(Tob1), (Tob5)	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM
2	Missing (lifetime use imputed)	Missing	0		
3	(Past month)	Missing	24	(Tob1), (Tob4), (Tob5), (Tob6)	1. D 2. PM
4	Not past year		313	(Tob1), (Tob3), (Tob5)	1. R3/(R3+R4)
5	Not past month		269	(Tob1), (Tob2), (Tob5)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
6	30-day frequency logically assigned based on estimated value, no missing values.		144	(Tob1), (Tob5)	
	Lifetime user, nothing missing		40072	(None)	
	Imputed to lifetime nonuse		0	(None)	
	Lifetime nonuser, nothing missing		27259	(None)	

[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 G.6 Restrictions and Portion of the Predictive Mean Vector for Cigar Users

Missingness Pattern					
#	Recency	30-Day Frequency	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	Past year	Missing	20	(Tob1)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
2	Missing (Lifetime use known)	Missing	22	(Tob1)	1. R1 2. R2 3. R3 4. R1*PM
2	Missing (Lifetime use imputed)	Missing	1		
3	(Past month)	Missing	11	(Tob1), (Tob4), (Tob6)	1. PM
4	Not past year		245	(Tob1), (Tob3)	1. R3/(R3+R4)
5	Not past month		286	(Tob1), (Tob2)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
6	30-day frequency logically assigned based on estimated value, no missing values.		36	(Tob1)	
	Lifetime user, nothing missing		22756		
	Imputed to lifetime nonuse		5		
	Lifetime nonuser, nothing missing		44780		

[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 G.7 Constraints for Smokeless Tobacco (Chewing Tobacco and Snuff)

Constraint #	Description
SLT1	If the difference between the recipient's current age and his or 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 or 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	If donor is not a chew user, then recipient must also not be a chew user (and vice versa)
SLT4	If donor is not a snuff user, then recipient must also not be a snuff user (and vice versa)
SLT5	If recipient's age at first chew use equals his or her current age, the donor's 30-day chew frequency (1) cannot be greater than the number of days between the recipient's interview date and his or her date of first chew use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his or her birthday (inclusive)
SLT6	If recipient's age at first snuff use equals his or her current age, the donor's 30-day snuff frequency (1) cannot be greater than the number of days between the recipient's interview date and his or her date of first snuff use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his or her birthday (inclusive)
SLT7	Donor must be a past month chew user (chew recency = 1)
SLT8	Donor must be a past month snuff user (snuff recency = 1)
SLT9	Donor's snuff recency 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 G.8 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco)

Missingness Pattern					Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.			
1	(Past month)	(Past month)	Missing	Missing	1	(SLT1-SLT4), (SLT5-SLT8)	1. DC 2. PMC 3. DS 4. PMS
2	(Past month)		Missing		1	(SLT1-SLT4), (SLT5), (SLT7), (SLT9)	1. DC 2. PMC
3		(Past month)		Missing[1]	5	(SLT1-SLT4), (SLT6), (SLT8), (SLT10)	1. DS 2. PMS
4		Missing (Lifetime use known)		Missing	5	(SLT1-SLT4), (SLT6), (SLT10)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
4		Missing (Lifetime use imputed)		Missing	3		
5	(Past month)	Missing (Lifetime use known)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT10)	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
5	(Past month)	Missing (Lifetime use imputed)	Missing	Missing	0		
6	Missing (lifetime use known)		Missing		4	(SLT1-SLT4), (SLT5), (SLT9)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
6	Missing (lifetime use imputed)		Missing		0		
7	Missing (lifetime use known)	(Past month)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT8)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS
7	Missing (lifetime use imputed)	(Past month)	Missing	Missing	0		

(continued)

Exhibit G.8 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	**Chew Recency**	**Snuff Recency**	**Chew 30-Day Freq.**	**Snuff 30-Day Freq.**	**Number of Cases**	**Logical Constraints**	**Predictive Mean Vector[1]**
8		Past year		Missing	4	(SLT1-SLT4), (SLT10-SLT11)	1. R1/(R1+R2) 2. RS1*DS/(RS1+RS2) 3. RS1*(1-DS)*PMS/(RS1+RS2)
9	Past year		Missing		5	(SLT1-SLT4), (SLT5), (SLT8), (SLT12)	1. R1/(R1+R2) 2. RC1*DC/(RC1+RC2) 3. RC1*(1-DC)*PMC/(RC1+RC2)
10	Missing (lifetime use known)	Missing (Lifetime use known)	Missing	Missing	1	(SLT1-SLT4), (SLT5-SLT6)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
10	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
10	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
10	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
11	Not past year				63	(SLT1-SLT4), (SLT8), (SLT13)	1. R3/(R3+R4)
12		Not past year			63	(SLT1-SLT4), (SLT10), (SLT14)	1. R3/(R3+R4)
13	Not past year	Not past year			14	(SLT1-SLT4), (SLT13-SLT14)	1. R3/(R3+R4)

(continued)

Exhibit G.8 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern					Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.			
14	Not past month				77	(SLT1-SLT4), (SLT9), (SLT15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
15		Not past month			98	(SLT1-SLT4), (SLT10), (SLT16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
16	Not past month	Not past month			13	(SLT1-SLT4), (SLT15-SLT16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
17	Not past month	(Past month)		Missing	0	(SLT1-SLT4), (SLT6), (SLT8), (SLT15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DS 4. PMS
18	(Past month)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT7), (SLT16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC
19	Not past month	Missing (lifetime use known)		Missing	0	(SLT1-SLT4), (SLT6), (SLT15)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
19	Not past month	Missing (lifetime use imputed)		Missing	0		
20	Missing (lifetime use known)	Not past month	Missing		1	(SLT1-SLT4), (SLT5), (SLT16)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
20	Missing (lifetime use imputed)	Not past month	Missing		0		
21	Not past month	Not past year			0	(SLT1-SLT4), (SLT14-SLT15)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. R3/(R3+R4)
22	Not past year	Not past month			0	(SLT1-SLT4), (SLT13), (SLT16)	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. R3/(R3+R4)
23	(Lifetime use of snuff, chewing tobacco, or both missing in raw data. Missing values imputed to nonuse in lifetime imputation; nothing missing at this point in sequence)				0		

(continued)

Exhibit G.8 Restrictions and Portion of the Predictive Mean Vector for Smokeless Tobacco Users (Snuff and Chewing Tobacco) (continued)

Missingness Pattern							
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
24	Not past year	Missing (lifetime use known)		Missing	0	(SLT1-SLT4), (SLT6), (SLT13)	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
24	Not past year	Missing (lifetime use imputed)		Missing	0		
25	Missing (lifetime use known)	Not past year	Missing		0	(SLT1-SLT4), (SLT5), (SLT14)	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
25	Missing (lifetime use imputed)	Not past year	Missing		0		
26	Past year	Past year	Missing	Missing	1	(SLT1-SLT4), (SLT5-SLT6), (SLT11-SLT12)	1. R1/(R1+R2) 2. RC1*(1-DC)*PMC 3. RC1*DC/ (RC1+RC2) 4. RC1*(1-DC)*PMC/ (RC1+RC2) 5. RS1*DS/ (RS1+RS2) 6. RS1*(1-DS)*PMS/ (RS1+RS2)
	Lifetime user, nothing missing				12517		
	Imputed to lifetime nonuse				39		
	Lifetime nonuser, nothing missing				55212		

[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 G.9 Pipe User Restrictions

Missingness Pattern			
#	Recency	Number of Cases	Constraints
1	Missing (lifetime use known)	2	(None)
1	Missing (lifetime use imputed)	2	(None)
	Lifetime user, nothing missing	6482	
	Imputed to lifetime nonuse	6	
	Lifetime nonuser, nothing missing	61634	

Note: For pipes, only a two-level recency-of-use variable was imputed. The imputation was univariate, both in terms of the predictive mean vector and the final assignment. Item nonrespondents were handled identically, whether or not lifetime use was imputed.

Exhibit G.10 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 or her current age, the donor's 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first drug use (+1) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (+1)

(continued)

Exhibit G.10 Constraints for Various Drugs (continued)

Drug	Constraint #	Constraint
Alc, Mrj, Inh	C12	If recipient's age at first use equals his or her current age, (1) recipient's donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (+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 or 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 or her current age, (1) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (-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 or her current age, (1) donor's DR5DAY must be less than recipient's days between his or her interview date and date of first drug use (+1) and (2) donor's DR5DAY must be less than recipient's days between his or 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 or her current age, (1) donor's proportion of past year used * recipient's max number of days could have used in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (-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 or 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 G.11 Restrictions and Portion of the Predictive Mean Vector for Alcohol Users

Missingness Pattern							
#	Recency	12-Month Freq.	30-Day Freq.	30-Day Binge Drink	Number of Cases	Logical Constraints	Predictive Mean Vector[1]
1	(Past month)	Missing	Missing		16	(C1-C13)	1. PM 2. PY
2	(Past month)		Missing		212	(C5-C8), (C10), (C11), C13	1. PM
3	(Past month)	Missing			152	(C1-C4), (C10), (C12)	1. PY
4	(Past year but not past month)	Missing			133	(C1-C3), (C14), (C15)	1. PY
5	(Past month)			Missing	553	(C10), (C16), (C17)	1. PMB
6	(Past month)		Missing	Missing	22	(C5-C7), (C10), (C11), (C13)	1. PM 2. PMB
7	(Past month)	Missing		Missing	45	(C1-C4), (C10), (C12), (C16), (C17)	1. PY 2. PMB
8	(Past month)	Missing	Missing	Missing	21	(C1-C4), (C5-C7), (C9-C13)	1. PM 2. PY 3. PMB
9	Past Year		Missing	Missing	388	(C5-C7), (C11), (C13, C15)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)
10	Past year	Missing	Missing	Missing	58	(C1-C3), (C5-C9), (C11-C14), (C18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)
11	Lifetime (known)	Missing	Missing	Missing	439	(C1-C7), (C9), (C11-C14)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB
11	Lifetime (imputed)	Missing	Missing	Missing	10	(C1-C70, (C9), (C11-C14)	
	(30-day binge drink response missing in raw data. Logically set to zero based on responses in other parts of questionnaire. No other responses missing.)				50		
	Lifetime user, nothing missing				47928		
	Imputed to lifetime nonuse				10		
	Lifetime nonuser, nothing missing				18139		

[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 G.12 Restrictions and Portion of the Predictive Mean Vector for Marijuana Users

Missingness Pattern				Number of Cases	Constraints	Predictive Mean Vector[1]
#	Recency	12-Month Freq.	30-Day Freq.			
1	(Past month)	Missing	Missing	9	(C1-C7), (C9-C13)	1. PM 2. PY
2	(Past month)		Missing	13	(C5-C7), (C10), (C11), (C13)	1. PM
3	(Past month)	Missing		57	(C1-C4), (C10), (C12)	1. PY
4	(Past year but not past month)	Missing		59	(C1-C3), (C13), (C14)	1. PY
5	Past year		Missing	115	(C5-C7), (C11), (C13), (C18)	1. R1/(R1+R2) 2. R1*PM/(R1*R2)
6	Past year	Missing	Missing	96	(C1-C3), (C5-C7), (C9), (C11-C14), (C18), (C19)	1. R1/(R1+R2) 2. R1*PM/(R1*R2) 3. PY
7	Missing (lifetime use known)	Missing	Missing	273	(C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20) (C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20)	1. R1 2. R2 3. R1*PM 3. (R1+R2)*PY
7	Missing (lifetime use imputed)	Missing	Missing	28		
	Lifetime user, nothing missing			27428		
	Imputed to lifetime nonuse			23		
	Lifetime nonuser, nothing missing			40025		

[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 G.13 Restrictions and Portion of the Predictive Mean Vector for Inhalant Users

Missingness Pattern				Number of Cases	Constraints	Predictive Mean Vector[1]
#	Recency	12-Month Freq.	30-Day Freq.			
1	(Past month)	Missing	Missing	7	(C1-C7), (C10), (13)	1. PM 2. PY
2	(Past month)		Missing	4	(C6-C8), (C10), (C13)	1. PM
3	(Past month)	Missing		14	(C1-C4), (C10)	1. PY
4	(Past year not past month)	Missing		20	(C1-C3), (C18)	1. PY
5	Past year		Missing	24	(C5-C7), (C9),(C13), (C18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	4	(C1-C3), (C5-C7), (C9), (C13), (C18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing (lifetime use known)	Missing	Missing	277	(C1-C3), (C5-C7), (C9), (C13)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
7	Missing (lifetime use imputed)	Missing	Missing	8	(C1-C3), (C5-C7), (C9), (C13)	
	Lifetime user, nothing missing			8294		
	Imputed to lifetime nonuse			107		
	Lifetime nonuser, nothing missing			59367		

[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 G.14 Restrictions and Portion of the Predictive Mean Vector for Heroin Users

Missingness Pattern				Number of Cases	Constraints	Predictive Mean Vector[1]
#	**Recency**	**12-Month Freq.**	**30-Day Freq.**			
1	(Past month)	Missing	Missing	1	(C1-C7), (C9), (C10-C13), (C21)	1. PM 2. PY
2	(Past month)		Missing	2	(C5-C7), (C10), (C13)	1. PM
3	(Past month)	Missing		1	(C1-C4), (C10), (C21)	1. PY
4	(Past year but not past month)	Missing		1	(C1-C3), (C15)	1. PY
5	Past year		Missing	2	(C5-C7), (C9), (C13), (C18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	8	(C1-C3), (C5-C7), (C9), (C13), (C18), (C21)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing (lifetime use known)	Missing	Missing	15	(C1-C3), (C5-C7), (C9), (C13), (C21) (C1-C3), (C5-C7), (C9), (C13), (C21)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
7	Missing (lifetime use imputed)	Missing	Missing	1		
	Lifetime user, nothing missing			871		
	Imputed to lifetime nonuse			40		
	Lifetime nonuser, nothing missing			67184		

[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 G.15 Restrictions and Portion of the Predictive Mean Vector for Users of Pain Relievers, Tranquilizers, and Sedatives

Missingness Pattern			Number of Cases	Constraints	Predictive Mean Vector[1]
#	Recency	12-Month Frequency			
1	(Past month)	Missing	Pain relievers: 34 Tranquilizers: 7 Sedatives: 3	(C1-C3), (C10)	1. PY
2	(Past year but not past month)	Missing	Pain relievers: 47 Tranquilizers: 9 Sedatives: 2	(C1-C3), (C15)	1. PY
3	Past year		Pain relievers: 5 Tranquilizers: 1 Sedatives: 0	(C18)	1. R1/(R1+R2)
4	Past year	Missing	Pain relievers: 19 Tranquilizers: 6 Sedatives: 1	(C1-C3), (C18)	1. R1/(R1+R2) 2. PY
5	Missing (lifetime use known)	Missing	Pain relievers: 327 Tranquilizers: 105 Sedatives: 34	(C1-C3), (C18) (C1-C3), (C18)	1. R1 2. R2 3. (R1+R2)*PY
5	Missing (lifetime use imputed)	Missing	Pain relievers: 26 Tranquilizers: 5 Sedatives: 1		
	Lifetime user, nothing missing		Pain relievers: 10036 Tranquilizers: 5194 Sedatives: 1791		
	Imputed to lifetime nonuse		Pain relievers: 225 Tranquilizers: 117 Sedatives: 142		
	Lifetime nonuser, nothing missing		Pain relievers: 57407 Tranquilizers: 62682 Sedatives: 66152		

Note: The missingness patterns and predictive mean vectors for the pain relievers, tranquilizers, and sedatives modules were identical.

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

Exhibit G.16 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 or her current age, the donor's cocaine 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first cocaine use (+1) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (+1)

(continued)

Exhibit G.16 Constraints for Cocaine and Crack (continued)

Constraint #	Constraint
Coc10	If recipient's age at first cocaine use equals his or her current age, (1) recipient's donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (+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 or 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 or her current age, donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year cannot be greater than recipient's days between his or her interview date and date of first cocaine use (-29)
Coc15	Donor must be a past month or past year (but not past month) cocaine user (cocaine recency = 1 or 2)
Coc16	Donor must be a past month, past year (but not past month), or a lifetime (but not past year) cocaine user (cocaine recency = 1, 2, or 3**)**
Coc17	If recipient's age at first cocaine use equals his or her current age, donor cannot be a lifetime (but not past year) cocaine user (cocaine recency cannot equal 3**)**
Coc18	Donor's proportion of past year crack use * recipient's max number of days could have used crack in past year must be less than (or equal) the recipient's maximum possible past year crack frequency of use. The recipient's maximum possible crack frequency of use in the past year is limited by the following factors: (1) it must be less 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

(continued)

Exhibit G.16 Constraints for Cocaine and Crack (continued)

Constraint #	Constraint
Coc20	(Recipient's proportion of past year crack use * max number of days could have used crack in past year) 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 or her current age, the donor's crack 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first crack use (+1) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (+1**)**
Coc26	If recipient's age at first crack use equals his or her current age, (1) recipient's donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (+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 or 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 or her current age donor cannot be a lifetime (but not past year) crack user (crack recency cannot equal 3**)**
Coc33	Donor must be a past year (but not past month) crack user (crack recency = 2)
Coc34	If recipient's age at first crack use equals his or her current age, donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than recipient's days between his or her interview date and date of first crack use (-29)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
1	(Past month)		Missing		Missing		8	(Coc1-Coc12)	1. PM 2. PY
2	(Past month)				Missing		14	(Coc1), (Coc6-Coc9), (Coc11-Coc12)	1. PM
3	(Past month)		Missing				7	(Coc2-Coc4), (Coc10), (Coc12)	1. PY
4	(Past year not past month)		Missing				42	(Coc2-Coc4), (Coc12-Coc14)	1. PY
5	Past year				Missing		25	(Coc6-Coc9), (Coc11-Coc12), (Coc15)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year		Missing		Missing		12	(Coc2-Coc12), (Coc15)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing (lifetime use known)		Missing		Missing		93	(Coc2-Coc12), (Coc16-Coc17)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
7	Missing (lifetime use imputed)		Missing		Missing		4		
8	(Past month)	(Past month)		Missing		Missing	0	(Coc1), (Coc18-Coc27)	1. PM 2. PY

(continued)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern									
#	**Cocaine Recency**	**Crack Recency**	**Cocaine 12-Mo. Freq.**	**Crack 12-Mo. Freq.**	**Cocaine 30-Day Freq.**	**Crack 30-Day Freq.**	**Number of Cases**	**Constraints**	**Predictive Mean Vector[1]**
9	(Past month)	(Past month)				Missing	0	(Coc1), (Coc22-Coc25), (Coc27-Coc28)	1. PM
10	(Past month)	(Past month)		Missing			2	(Coc15), (Coc18-Coc20), (Coc26), (Coc28)	1. PM
11	(Past year not missing)	(Past year not past month)		Missing			0	(Coc15), (Coc18-Coc20), (Coc26), (Coc29)	1. PY
12	(Past month)	Past year				Missing	3	(Coc1), (Coc22-Coc25), (Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
13	(Past month)	Past year		Missing		Missing	1	(Coc1), (Coc18-Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
14	(Past month)	Missing (Lifetime use known)		Missing		Missing	6	(Coc16), (Coc18-Coc26), (Coc30-Coc32)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
14	(Past month)	Missing (Lifetime use imputed)		Missing		Missing	0		
15	(Past month)	(Past month)	Missing	Missing			0	(Coc1-Coc4)), (Coc10), (Coc18-Coc20), (Coc26), (Coc28)	1. PM

(continued)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.			
16	(Past month)	(Past year but not past month)	Missing	Missing			0	(Coc1-Coc4), (Coc10), (Coc18-Coc20), (Coc26), (Coc33)	1. PY
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			1	(Coc2-Coc4), (Coc14), (Coc18-Coc20), (Coc33-Coc34)	1. PY
18	(Past month)	(Past month)			Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc22-Coc25), (Coc27-Coc28)	1. PM
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	1	(Coc1-Coc11), (Coc18-Coc28)	1. PM 2. PY
20	(Past month)	(Past month)	Missing		Missing	Missing	0	(Coc1-Coc11), (Coc16), (Coc22-Coc25), (Coc27-Coc28)	1. PM
21	(Past month)	(Past month)		Missing	Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc28)	1. PM
22	(Past month)	(Past month)	Missing	Missing	Missing		0	(Coc1-Coc11), (Coc18-Coc21), (Coc26), (Coc28)	1. PM 2. PY
23	(Past month)	(Past month not past year)	Missing	Missing	Missing		0	(Coc1-Coc11), (Coc18-Coc20), (Coc33), (Coc34)	1. PM 2. PY

(continued)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.			
24	(Past month)	(Past month)	Missing	Missing		Missing	0	(Coc1-Coc4), (Coc10), (Coc18-Coc26), (Coc28)	1. PM
25	(Past month)	(Past month)		Missing	Missing		1	(Coc1), (Coc6-Coc9), (Coc18-Coc20), (Coc26), (Coc28)	1. PM
26	(Past month)	(Past year not past month)		Missing	Missing		0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc 20), (Coc26), (Coc33)	1. PY
27	(Past month)	(Past month)	Missing			Missing	0	(Coc1-Coc4), (Coc10), (Coc22-Coc25), (Coc27-Coc28)	1. PM
28	Past year	Past year			Missing	Missing	3	(Coc6-Coc9), (Coc11), (Coc15), (Coc22-Coc25), (Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
29	Past year	Past year	Missing		Missing	Missing	1	(Coc3-Coc11), (Coc15), (Coc21-Coc25), (Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
30	Past year	Past year		Missing	Missing	Missing	5	(Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

(continued)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.			
31	Past year	Past year	Missing	Missing	Missing	Missing	1	(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc29)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
32	Past year	Missing (lifetime use known)		Missing	Missing	Missing	2	(Coc1), (Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc30)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
32	Past year	Missing (lifetime use imputed)		Missing	Missing	Missing	0		
33	Past year	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc30), (Coc32)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
33	Past year	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0		
34	(Past month)	Missing (lifetime use known)		Missing	Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc27), (Coc30), (Coc32)	1. PM 2. PY
34	(Past month)	Missing (lifetime use imputed)		Missing	Missing	Missing	0		
	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	(Coc1-Coc11), (Coc18-Coc27), (Coc30)	1. PM 2. PY
35	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0		

(continued)

Exhibit G.17 Restrictions and Portion of the Predictive Mean Vector for Cocaine Users (continued)

Missingness Pattern									
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	20	(Coc2-Coc11), (Coc16-Coc27), (Coc30)	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
36	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0		
36	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0		
36	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	1		
	Lifetime user, nothing missing						7982		
	Imputed to lifetime nonuse						30		
	Lifetime nonuser, nothing missing						59882		

Note: Included crack users and cocaine users who were not crack users.

[1] The predictive mean vector components are defined by the following: 1. R1 = P(past month cocaine use | lifetime cocaine use). 2. R2 = P(past year but not past month cocaine use | lifetime cocaine use). 3. PM = P(cocaine use on a given day in the past month | past month use of cocaine). 4. PY = P(cocaine use on a given day in the past year | past year use of cocaine).

Exhibit G.18 Constraints for Hallucinogens (Including LSD, PCP, and ECS)

Constraint #	Constraint
Hal1	Donor's proportion of past year hallucinogen use * recipient's max number of days could have used hallucinogens in past year must be less than (or equal) 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)
Hal2	Donor's proportion of past year hallucinogen use * recipient's min number of days could have used hallucinogens in past year must be greater than (or equal) the recipient's minimum possible past year hallucinogen frequency of use. The recipient's minimum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month hallucinogen user, it must be at least as much as the hallucinogen 30-day freq (2) if the recipient is not a past month hallucinogen user but a past year hallucinogen user, it must be at least 1
Hal3	(Recipient's proportion of past year hallucinogen use * max number of days could have used hallucinogens in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Hal4	Donor's 30-day hallucinogen use less than number of days between recipient's interview date and birthday (+1)
Hal5	Donor's 30-day hallucinogen use less than the recipient's maximum number of days could have used hallucinogens in past 30 days
Hal6	Donor's 30-day hallucinogen use greater than the recipient's minimum number of days could have used hallucinogens in past 30 days
Hal7	Donor must be a LSD user (LSD recency not equal to 91)
Hal8	Donor must be a PCP user (PCP recency not equal to 91)
Hal9	Donor must be an ECS user (ECS recency not equal to 91)
Hal10	Donor's LSD recency must equal recipient's LSD recency
Hal11	Donor's PCP recency must equal recipient's PCP recency
Hal12	Donor's ECS recency must equal recipient's ECS recency
Hal13	Donor must be a LSD and PCP user (LSD and PCP recencies not equal to 91)

(continued)

Exhibit G.18 Constraints for Hallucinogens (Including LSD, PCP, and ECS) (continued)

Constraint #	Constraint
Hal14	Donor must be a LSD and ECS user (LSD and ECS recencies not equal to 91)
Hal15	Donor must be a PCP and ECS user (PCP and ECS recencies not equal to 91)
Hal16	Donor must be a LSD and PCP and ECS user (LSD and PCP and ECS recencies not equal to 91)
Hal17	Donor's must be a past month hallucinogens user (hallucinogen recency = 1)
Hal18	Donor must be a hallucinogen past year (but not past month) or past month user (hallucinogen recency = 1 or 2)
Hal19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
Hal20	Donor must be a LSD past year (but not past month) or past month user (LSD recency = 1 or 2)
Hal21	Donor must be a PCP past year (but not past month) or past month user (PCP recency = 1 or 2)
Hal22	Donor must be an ECS past year (but not past month) or past month user (ECS recency = 1 or 2)
Hal23	Donor must not be a LSD past year (but not past month) or past month user (LSD recency not equal to 1 or 2)
Hal24	Donor must not be a PCP past year (but not past month) or past month user (PCP recency not equal to 1 or 2)
Hal25	Donor must not be an ECS past year (but not past month) or past month user (ECS recency not equal to 1 or 2)
Hal26	Donor's hallucinogen recency must equal recipient's hallucinogen recency or donor's hallucinogen recency must equal recipient's hallucinogen recency minus 10

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
1		Missing (lifetime use known)					2	(Hal7,11,12,26)	1. R1 2. R2
1		Missing (lifetime use imputed)					0		
2			Missing (lifetime use known)				1	(Hal8,10,12,26)	1. R1 2. R2
2			Missing (lifetime use imputed)				0		
3		Missing (lifetime use known)	Missing (lifetime use known)				0	(Hal7,8,12,26)	1. R1 2. R2
3		Missing (lifetime use known)	Missing (lifetime use imputed)				0		
3		Missing (lifetime use imputed)	Missing (lifetime use known)				0		
3		Missing (lifetime use imputed)	Missing (lifetime use imputed)				0		
4	(Past month)				Missing	Missing	66	(Hal1-6,17)	1. PM 2. PY
5	(Past month)					Missing	75	(Hal4-6,17)	1. PM
6	(Past year)				Missing		243	(Hal1-3,18)	1. PY
7	(Past month)	Missing (lifetime use known)				Missing	3	(Hal4-6,7,11,12,17)	1. R1 2. R2 3. PM

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
7	(Past month)	Missing (lifetime use imputed)				Missing	0		
8	(Past month)		Missing (lifetime use known)			Missing	1	(Hal4-6,8,10,12,17)	1. R1 2. R2 3. PM
8	(Past month)		Missing (lifetime use imputed)			Missing	0		
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)			Missing	0	(Hal4-6,7,8,12,17)	1. R1 2. R2 3. PM
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)			Missing	0		
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)			Missing	0		
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			Missing	0		
10	(Past month or Past month not past year)	Missing (lifetime use known)			Missing		0	(Hal1-3,7,11,12,18)	1. R1 2. R2 3. PY
10	(Past month or Past month not past year)	Missing (lifetime use imputed)			Missing		0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
11	(Past month or Past month not past year)		Missing (lifetime use known)		Missing		0	(Hal1-3,8,10,12,18)	1. R1 2. R2 3. PY
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	(Hal1-3,7,8,12,18)	1. R1 2. R2 3. PY
12	Past year (not missing)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing		0		
12	Past year (not missing)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing		0		
12	Past year (not missing)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing		0		
13	(Past month)	Missing (lifetime use known)			Missing	Missing	0	(Hal1-6,7,11,12,17)	1. R1 2. R2 3. PM 4. PY
13	(Past month)	Missing (lifetime use imputed)			Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
14	(Past month)		Missing (lifetime use known)		Missing	Missing	1	(Hal1-6,8,10,12,17)	1. R1 2. R2 3. PM 4. PY
14	(Past month)		Missing (lifetime use imputed)		Missing	Missing	0		
15	(Past month	Missing (lifetime use known)	Missing (lifetime use known)		Missing	Missing	3	(Hal1-6,7,8,12,17)	1. R1 2. R2 3. PM 4. PY
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	Missing	0		
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	Missing	0		
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	Missing	0		
16	Past year	(Not past month)	(Not past month)	(Not past month)		Missing	20	(Hal4-6,10-12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
17	Past year	(Not past month)	(Not past month)	(Not past month)	Missing	Missing	7	(Hal1-6,10-12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
18	Past year	Past year	(Not past month)	(Not past month)		Missing	6	(Hal4-6,11,12,18,20)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
19	Past year	(Not past month)	Past year	(Not past month)		Missing	5	(Hal4-6,10,12,18,21)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
20	Past year	Past year	Past year	(Not past month)		Missing	2	(Hal4-6,12,18,20,21)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
21	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)		Missing	14	(Hal4-6,7,11,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
21	Past year	Missing (lifetime use imputed)	(Not past month)	(Not past month)		Missing	0		
22	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)		Missing	7	(Hal4-6,8,10,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
22	Past year	(Not past month)	Missing (lifetime use imputed)	(Not past month)		Missing	0		
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)		Missing	0	(Hal4-6,7,8,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
23	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past month)		Missing	0		
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past month)		Missing	0		
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past month)		Missing	0		
24	Past year	Past year	(Not past month)	(Not past month)	Missing	Missing	3	(Hal1-6,11,12,18,20)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
25	Past year	(Not past month)	Past year	(Not past month)	Missing	Missing	1	(Hal1-6,10,12,18,21)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
26	Past year	Past year	Past year	(Not past month)	Missing	Missing	0	(Hal1-6,,12,18,20,21)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
27	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)	Missing	Missing	3	(Hal1-6,7,11,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
27	Past year	Missing (lifetime use imputed)	(Not past month)	(Not past month)	Missing	Missing	0		
28	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)	Missing	Missing	4	(Hal1-6,8,11,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
28	Past year	(Not past month)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0		
29	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0	(Hal1-6,7,8,12,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
29	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0		
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0		
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
30	Missing (lifetime use known)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	43	(Hal1-6,10-12,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
30	Missing (lifetime use imputed)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	5		
31	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	59	(Hal1-6,7,11,12,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
31	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	0		
31	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	0		
31	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	1		
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	13	(Hal1-6,8,10,12,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	**Hallucinogen Recency**	**LSD Recency**	**PCP Recency**	**ECS Recency**	**Hallucinogen 12-Mo. Freq.**	**Hallucinogen 30-Day Freq.**	**Number of Cases**	**Constraints**	**Predictive Mean Vector[1]**
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0		
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	5	(Hal1-6,7,8,12,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0		
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	1		
34				Missing (lifetime use known)			3	(Hal9-11,26)	1. R1 2. R2
34				Missing (lifetime use imputed)			0		
35		Missing (lifetime use known)		Missing (lifetime use known)			2	(Hal7,9,11,26)	1. R1 2. R2
35		Missing (lifetime use known)		Missing (lifetime use imputed)			0		
35		Missing (lifetime use imputed)		Missing (lifetime use known)			0		
35		Missing (lifetime use imputed)		Missing (lifetime use imputed)			0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
36			Missing (lifetime use known)	Missing (lifetime use known)			0	(Hal8,9,10,26)	1. R1 2. R2
36			Missing (lifetime use known)	Missing (lifetime use imputed)			0		
36			Missing (lifetime use imputed)	Missing (lifetime use known)			0		
36			Missing (lifetime use imputed)	Missing (lifetime use imputed)			0		
37		Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)			0	(Hal7-9,26)	1. R1 2. R2
37		Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)			0		
37		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)			0		
37		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
37		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)			0		
37		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)			0		
37		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)			0		
37		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)			0		
38	(Past month)			Missing (lifetime use known)		Missing	0	(Hal4-6,9,10,11,17)	1. R1 2. R2 3. PM
38	(Past month)			Missing (lifetime use imputed)		Missing	0		
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)		Missing	0	(Hal4-6,7,9,11,17)	1. R1 2. R2 3. PM
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use imputed)		Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
39	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use known)		Missing	0		
39	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use imputed)		Missing	0		
40	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)		Missing	1	(Hal4-6,8,9,10,17)	1. R1 2. R2 3. PM
40	(Past month)		Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		
40	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	(Hal4-6,7,8,9,17)	1. R1 2. R2 3. PM
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0		
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
41	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
42	(Past year)			Missing (lifetime use known)	Missing		0	(Hal1-3,9,10,11,18)	1. R1 2. R2 3. PY
42	(Past year)			Missing (lifetime use imputed)	Missing		0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use known)	Missing		0	(Hal1-3,7,9,11,18)	1. R1 2. R2 3. PY
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use imputed)	Missing		0		
43	(Past year)	Missing (lifetime use imputed)		Missing (lifetime use known)	Missing		0		
43	(Past year)	Missing (lifetime use imputed)		Missing (lifetime use imputed)	Missing		0		
44	(Past year)		Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	(Hal1-3,8,9,10,18)	1. R1 2. R2 3. PY
44	(Past year)		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0		
44	(Past year)		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0		
44	(Past year)		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	(Hal1-3,7,8,9,18)	1. R1 2. R2 3. PY
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0		
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0		
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0		
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0		
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0		
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0		
45	(Past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
46	(Past month)			Missing (lifetime use known)	Missing	Missing	1	(Hal1-6,9,10,11,17)	1. R1 2. R2 3. PM 4. PY
46	(Past month)			Missing (lifetime use imputed)	Missing	Missing	0		
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,9,11,17)	1. R1 2. R2 3. PM 4. PY
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use imputed)	Missing	Missing	0		
47	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use known)	Missing	Missing	0		
47	(Past month)	Missing (lifetime use imputed)		Missing (lifetime use imputed)	Missing	Missing	0		
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,8,9,10,17)	1. R1 2. R2 3. PM 4. PY
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
48	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
48	(Past month)		Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,8,9,17)	1. R1 2. R2 3. PM 4. PY
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
49	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
50	Past year	(Not past month)	(Not past month)	Past year		Missing	9	(Hal4-6,10,11,18,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
51	Past year	Past year	(Not past month)	Past year		Missing	1	(Hal4-6,11,18,20,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
52	Past year	(Not past month)	Past year	Past year		Missing	2	(Hal4-6,10,18,21,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
53	Past year	Past year	Past year	Past year		Missing	0	(Hal4-6,18,20-22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)		Missing	19	(Hal4-6,9,10,11,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)		Missing	0		
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)		Missing	1	(Hal4-6,7,9,11,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use imputed)		Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.			
55	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use known)		Missing	0		
55	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use imputed)		Missing	0		
56	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	1	(Hal4-6,8,9,10,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
56	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		
56	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
56	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	(Hal4-6,7,8,9,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
57	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
57	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0		
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0		
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0		
57	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0		
58	Past year	(Not past month)	(Not past month)	Past year	Missing	Missing	1	(Hal1-6,10,11,18,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
59	Past year	Past year	(Not past month)	Past year	Missing	Missing	0	(Hal1-6,11,18,20,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
60	Past year	(Not past month)	Past year	Past year	Missing	Missing	0	(Hal1-6,10,18,21,22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
61	Past year	Past year	Past year	Past year	Missing	Missing	1	(Hal1-6,18,20-22)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,9,10,11,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0		
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,9,11,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0		
63	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0		
63	Past year	Missing (lifetime use imputed)	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0		
64	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,8,9,10,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
64	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
64	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
64	Past year	(Not past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,8,9,18)	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
65	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
65	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
65	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	132	(Hal1-6,9,10,11,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
66	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0		
66	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	1		
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	9	(Hal1-6,7,9,11,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
67	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0		
67	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
67	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0		
67	Missing (lifetime use imputed)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
67	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing	Missing	0		
67	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0		
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	6	(Hal1-6,8,9,10,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
68	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,8,9,19)	1. R1 2. R2 3. R1*PM 4. (R1+R2)* PY

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued),

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
69	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0		
70				Past year			1	(Hal10,11,22,26)	1. R1/(R1+R2)
71		Past year	Past year				0	(Hal12,20,21,26)	1. R1/(R1+R2)
72		Past year		Past year			0	(Hal11,20,22,26)	1. R1/(R1+R2)

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern							Number of Cases	Constraints	Predictive Mean Vector[1]
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.			
73	(Past month)	Past year			Missing	Missing	1	(Hal1-6,11,12,17,20)	1. R1/(R1+R2) 2. PM 3. PY
74	(Past month)		Past year		Missing	Missing	0	(Hal1-6,10,12,17,21)	1. R1/(R1+R2) 2. PM 3. PY
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,7,9,18,21)	1. R1/(R1+R2) 2. PM 3. PY
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use imputed)	Missing	Missing	0		
75	Past year	Missing (lifetime use imputed)	Past year	Missing (lifetime use known)	Missing	Missing	0		
75	Past year	Missing (lifetime use imputed)	Past year	Missing (lifetime use imputed)	Missing	Missing	0		
76	Past year	(Not past month)	Past year	Missing (lifetime use known)	Missing	Missing	0	(Hal1-6,9,10,18,21)	1. R1/(R1+R2) 2. PM 3. PY
76	Past year	(Not past month)	Past year	Missing (lifetime use imputed)	Missing	Missing	0		
77			Past year				0	(Hal10,12,21,26)	1. R1/(R1+R2)
78		Past year					2	(Hal11,12,20,26)	1. R1/(R1+R2)

(continued)

Exhibit G.19 Restrictions and Portion of the Predictive Mean Vector for Hallucinogen Users (Including LSD, PCP, and ECS) (continued)

Missingness Pattern									
#	Hallucinogen Recency	LSD Recency	PCP Recency	ECS Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
79	(Past month)	Past year				Missing	1	(Hal4-6,7,11,12,17)	1. R1/(R1+R2) 2. PM
80	(Past month)		Past year			Missing	2	(Hal4-6,8,10,12,17)	1. R1/(R1+R2) 2. PM
81	(Past month)			Past year		Missing	1	(Hal4-6,9,10,11,17)	1. R1/(R1+R2) 2. PM
82	(Past month)	Past year			Missing		1	(Hal1-3,7,11,12,17)	1. R1/(R1+R2) 2. PY
83	(Past month)		Past year		Missing		0	(Hal1-3,8,10,12,17)	1. R1/(R1+R2) 2. PY
84	(Past month)			Past year	Missing		0	(Hal1-3,9,10,11,17)	1. R1/(R1+R2) 2. PY
	Lifetime user, nothing missing						10123		
	Imputed to lifetime nonuse						185		
	Lifetime nonuser, nothing missing						57127		

Note: Hallucinogen users included users of LSD, users of PCP, and users of ECS.

[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 G.20 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 methamphetamines recency equals the recipient's methamphetamines recency
Stm6	If recipient's age at first stimulants use equals his or 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 or 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 or 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 or 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 or 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 or 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)

(continued)

Exhibit G.20 Constraints for Stimulants and Methamphetamines (continued)

Constraint #	Constraint
Stm10	If recipient's age at first stimulants use equals his or her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first stimulants use (+1) and (2) cannot be greater than the recipient's days between his or 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) methamphetamines user (methamphetamines recency = 1, 2, or 3)
Stm13	If the number of days between the recipient's interview and birthday (+1) is between 0 and 30, methamphetamines 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, methamphetamines recency must not equal 3
Stm15	If recipient's age at first stimulants use equals his or her current age or the recipient's age at first methamphetamines use equals his or her current age or the recipient's number of days between his or her interview date and date at first methamphetamines use less than 30, the donor's recency must not equal 3
Stm16	If recipient's age at first stimulants use equals his or her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first stimulants use (-29) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (-29)
Stm17	Donor must be a past month or past year (but not past month) methamphetamines user (methamphetamines recency = 1 or 2)
Stm18	Donor's proportion of past year methamphetamines use * recipient's max number of days could have used methamphetamines in past year must be less than (or equal) the recipient's maximum possible past year methamphetamines frequency of use. The recipient's maximum possible methamphetamines 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 methamphetamines, as determined by the month of first use (2) if the maximum period the recipient could have used methamphetamines is greater than 30, but the recipient is a past month methamphetamines user with a nonmissing 30-day frequency, the past year methamphetamines 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 methamphetamines month user, the past year methamphetamines frequency must be less than or equal to the maximum period (30)

(continued)

Exhibit G.20 Constraints for Stimulants and Methamphetamines (continued)

Constraint #	Constraint
Stm19	Donor's proportion of past year methamphetamines use * recipient's min number of days could have used methamphetamines in past year must be greater than (or equal) the recipient's minimum possible past year methamphetamines frequency of use. The recipient's minimum possible methamphetamines frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month methamphetamines user, it must be at least as much as the 30-day freq (2) if the recipient is not a past month methamphetamines user but a past year methamphetamines user, it must be at least 1.
Stm20	(Recipient's proportion of past year methamphetamines use * max number of days could have used methamphetamines in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Stm21	If recipient's age at first methamphetamines use equals his or her current age, (1) recipient's donor's proportion of past year methamphetamines use * recipient's max number of days could have used methamphetamines in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (+1) and (2) donor's proportion of past year methamphetamines use * recipient's max number of days could have used methamphetamines in past year cannot be greater than the recipient's days between his or her interview date and birthday (+1)
Stm22	If recipient's age at first methamphetamines use equals his or her current age, (1) recipient's donor's proportion of past year methamphetamines use * recipient's max number of days could have used methamphetamines in past year cannot be greater than recipient's days between his or her interview date and date of first drug use (-29) and (2) donor's proportion of past year methamphetamines use * recipient's max number of days could have used methamphetamines in past year cannot be greater than the recipient's days between his or her interview date and birthday (-29)
Stm23	Donor must be a past month methamphetamines user (methamphetamines recency = 1)
Stm24	Donor must be a past year (but not past month) methamphetamines user (methamphetamines recency = 2)
Stm25	If recipient's age at first methamphetamines use equals his or her current age, the donor's methamphetamines 30-day frequency (1) cannot be greater than the recipient's days between his or her interview date and date of first methamphetamines use (+1) and (2) cannot be greater than the recipient's days between his or her interview date and birthday (+1)
Stm26	Donor must be a past month, past year (but not past month), or lifetime (but not past year) stimulants user (methamphetamines recency = 1, 2, or 3)

Exhibit G.21 Restrictions and Portion of the Predictive Mean Vector for Stimulant Users (Including Methamphetamines)

Missingness Pattern							
#	Stimulants Recency	Meth. Recency	Stimulants 12-Month Freq.	Meth. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
1	(Past month)		Missing		33	(Stm1-Stm6)	1. PY
2	(Past year but not past month)		Missing		52	(Stm1-Stm3), (Stm5), (Stm7-Stm8)	1. PY
3	Past year				2	(Stm5), (Stm8-Stm10)	1. R1/(R1+R2)
4	Past year		Missing		5	(Stm1-Stm3), (Stm5-Stm6), (Stm8-Stm9)	1. R1/(R1+R2) 2. PY
5	Missing (lifetime use known)		Missing		105	(Stm1-Stm3), (Stm5-Stm6), (Stm8)	1. R1 2. R2 3. (R1+R2)*PY
5	Missing (lifetime use imputed)		Missing		3		
6	(Past month)	(Past month)		Missing	0	(Stm4,Stm18-Stm23)	PY
7	(Past year not missing)	(Past year not past month)		Missing	1	(Stm9,Stm17-Stm23)	PY
8	(Past year not missing)	Past year			0	(Stm5,Stm8-Stm10)	1. R1/(R1+R2)
9	(Past year not missing)	Past year	Missing		0	(Stm1-Stm3),Stm5, (Stm8-Stm10)	1. R1/(R1+R2) 2. PY
10	(Past year not missing)	Past year		Missing	1	Stm5, (Stm8-Stm10), (Stm18-Stm20)	1. R1/(R1+R2) 2. PY
11	(Past year not missing)	Past year	Missing	Missing	0	(Stm1-Stm3,Stm5, Stm8-10, Stm18-Stm20)	1. R1/(R1+R2) 2. PY
12	(Past year not missing)	Missing (lifetime use known)		Missing	2	Stm5, (Stm8-Stm10), (Stm18-Stm20)	1. R1 2. R2 3. (R1+R2)*PY
12	(Past year not missing)	Missing (lifetime use imputed)		Missing	2		

(continued)

Exhibit G.21 Restrictions and Portion of the Predictive Mean Vector for Stimulant Users (Including Methamphetamines) (continued)

Missingness Pattern					Number of Cases	Constraints	Predictive Mean Vector[1]
#	Stimulants Recency	Meth. Recency	Stimulants 12-Month Freq.	Meth. 12-Month Freq.			
13	(Past month)	(Past month)	Missing	Missing	2	(Stm1-Stm3, Stm4, Stm23, Stm8, Stm10, Stm18-Stm20)	PY
14	(Past month)	(Past year not past month)	Missing	Missing	0	(Stm1-Stm3, Stm4, Stm24, Stm8, Stm10, Stm18-Stm20)	PY
15	(Past year not past month)	(Past year not past month)	Missing	Missing	6	(Stm1-Stm3, Stm7, Stm24, Stm8, Stm10, Stm18-Stm20)	PY
16	Past year	Past year			2	(Stm8-Stm10, Stm17, Stm22, Stm25)	R1/(R1+R2)
17	Past year	Past year	Missing		0	(Stm1-Stm3, Stm8-Stm10, Stm17, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
18	Past year	Past year		Missing	5	(Stm8-Stm10, Stm17-Stm20, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
19	Past year	Past year	Missing	Missing	9	(Stm1-Stm3, Stm8-Stm10, Stm17-Stm20, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
20	Past year	Missing (lifetime use known)		Missing	7	(Stm8-Stm10, Stm12, Stm18-Stm20, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
20	Past year	Missing (lifetime use imputed)		Missing	0		
21	Past year	Missing (lifetime use known)	Missing	Missing	2	(Stm1-Stm3, Stm8-Stm10, Stm12, Stm18-Stm20, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
21	Past year	Missing (lifetime use imputed)	Missing	Missing	0		
22	(Past month)	Missing (lifetime use known)		Missing	0	(Stm4, Stm8,Stm10, Stm12, Stm18-Stm20, Stm22, Stm25)	1. R1 2. R2 3. (R1+R2)*PY

(continued)

Exhibit G.21 Restrictions and Portion of the Predictive Mean Vector for Stimulant Users (Including Methamphetamines) (continued)

Missingness Pattern							
#	Stimulants Recency	Meth. Recency	Stimulants 12-Month Freq.	Meth. 12-Month Freq.	Number of Cases	Constraints	Predictive Mean Vector[1]
22	(Past month)	Missing (lifetime use imputed)		Missing	0		
23	(Past month)	Missing (lifetime use known)	Missing	Missing	1	(Stm1-Stm3, Stm4, Stm8,Stm10, Stm12, Stm18-Stm20, Stm22, Stm25)	1. R1 2. R2 3. (R1+R2)*PY
23	(Past month)	Missing (lifetime use imputed)	Missing	Missing	0		
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	46	(Stm1-Stm3, Stm5, Stm8,Stm10, Stm12, Stm18-Stm20, Stm22, Stm25-Stm26)	1. R1 2. R2 3. (R1+R2)*PY
24	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0		
24	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0		
24	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	2		
25	Past year	(Past year not past month)	Missing	Missing	1	(Stm1-Stm3, Stm8-Stm10, Stm17-Stm20, Stm22, Stm25)	1. R1/(R1+R2) 2. PY
	Lifetime user, nothing missing				5659		
	Imputed to lifetime nonuse				122		
	Lifetime nonuser, nothing missing				62104		

Note: Users of stimulants included users of methamphetamines.
1 The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

G.2.4 Source of Income

There were a large number of missingness patterns for the source of income variables because they were imputed simultaneously in a set. The only logical constraint applied to the potential donors was that they have the same value as the recipient for the imputation-revised family skip variable (IRFAMSKP). This logical constraint was applied for all missingness patterns.

Exhibit G.22 Restrictions and Portion of the Predictive Mean Vector for Income

<table>
<tr><th colspan="4">Missingness Pattern</th><th rowspan="2">Number of Cases</th><th rowspan="2">Constraints</th><th rowspan="2">Predictive Mean Vector[1]</th></tr>
<tr><th>#</th><th>Welfare Months</th><th>Family Payment</th><th>Family Service</th></tr>
<tr><td>1</td><td>missing</td><td>receiving</td><td>not receiving</td><td rowspan="3">163</td><td rowspan="6">irfamskp of donor should equal to that of recipient</td><td rowspan="3">WMS, and probabilities associated with other missing elements</td></tr>
<tr><td>2</td><td>missing</td><td>not receiving</td><td>receiving</td></tr>
<tr><td>3</td><td>missing</td><td>receiving</td><td>receiving</td></tr>
<tr><td>4</td><td>missing</td><td>not receiving</td><td>missing</td><td>103</td><td>SVC*WMS, SVC, and probabilities associated with other missing elements</td></tr>
<tr><td>5</td><td>missing</td><td>missing</td><td>not receiving</td><td>143</td><td>PMT*WMS, PMT, and probabilities associated with other missing elements</td></tr>
<tr><td>6</td><td>missing</td><td>missing</td><td>missing</td><td>381</td><td>[1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements</td></tr>
</table>

[1] The predictive mean vector components are defined by the following:
1. PMT = P(family in household received income from welfare payments)
2. SVC = P(family in household received income from other welfare services)
3. WMS = P(family in household received any welfare on a given month in the past year | family received any welfare in the past year)

G.2.5 Health Insurance

Both of the methods that were used to create the final imputation-revised health insurance variables, the “Old Method” and the “Constituent Variables Method,” are given in this section (see Chapter 10 for details).

G.2.5.1 Health Insurance (Old Method)

The health insurance variables IRINSUR (overall health insurance using only questions available in 1999 questionnaire), IRINSUR3 (overall health insurance using all questions available in 2001 and 2002 questionnaires), and IRPINSUR (private health insurance) were imputed as a set. Their edited counterparts were INSUR, INSUR3, and PINSUR. Details are in Chapter 10.

Exhibit G.23 Constraints for Health Insurance (Old Method)

Constraint #	Logical Constraint
HI2001_1	Donor must not have received private health insurance (PINSUR=0)[1]
HI2001_2	Donor must not have received overall health insurance by the 1999 definition (INSUR=0)
HI2001_3	Donor must have received overall health insurance by the 2001 definition (INSUR3=1)
HI2001_4	Donor must have received overall health insurance by the 1999 definition (INSUR=1)[1]

[1]Technically, these were not logical constraints. See Chapter 7 for details.

Exhibit G.24 Health Insurance (Old Method)

Missingness Pattern				Number of Cases	Logical Constraints	Predictive Mean Vector[1]
#	INSUR3	INSUR	PINSUR			
1	Missing	No	No	49	HI2001_1, HI2001_2	(OVR*(1-PRV))/(1-OVR*PRV)
2	Yes	Missing	No	211	HI2001_1, HI2001_3	(OVR*(1-PRV))/(1-OVR*PRV)
3	Missing	Missing	No	58	HI2001_1	(OVR*(1-PRV))/(1-OVR*PRV)
4	Yes	Missing	Missing	0	HI2001_3	OVR, OVR*PRV
5	Missing	Missing	Missing	0		OVR, OVR*PRV
6	Yes	Yes	Missing	0	HI2001_4	PRV

[1] The predictive mean vector components are defined by the following:
1. OVR = P(respondent received health insurance, 2001 definition)
2. PRV = P(respondent received private health insurance | respondent received health insurance, 2001 definition)

G.2.5.2 Health Insurance (Constituent Variables Method)

The health insurance variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT were imputed as a set. Their edited counterparts were CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Details are given in Chapter 10. The “Predictive Mean Vector” column is omitted from Exhibit G.25 because the elements of the vector were simply the predictive means associated with all missing variables. For example, for all missingness patterns where CAIDCHIP was missing, the probability that the respondent had CAIDCHIP=1 was included in the predictive mean vector. The “Logical Constraints” column is also omitted from Exhibit G.25 because no logical constraints were applied.

Exhibit G.25 Health Insurance (Constituent Variables Method)

Missingness Pattern					Number of Cases
#	CAIDCHIP	MEDICARE	CHAMPUS	PRVHLTIN	
1	Missing				106
2		Missing			30
3	Missing	Missing			13
4			Missing		30
5	Missing		Missing		23
6		Missing	Missing		4
7	Missing	Missing	Missing		11
8				Missing	114
9	Missing			Missing	29
10		Missing		Missing	3
11	Missing	Missing		Missing	11
12			Missing	Missing	9
13	Missing		Missing	Missing	40
14		Missing	Missing	Missing	1
15	Missing	Missing	Missing	Missing	67

Appendix H: Quality Control Measures Used in the Imputation Procedures

Appendix H: Quality Control Measures Used in the Imputation Procedures

H.1 Introduction

For the 2002 National Survey of Drug Use and Health (NSDUH),[143] the quality control (QC) imputation procedures as applied to demographic, drug use, income, health insurance and household composition (roster) variables are discussed in this Appendix. The imputation process occurred in three basic steps: (1) weight adjustment for item nonresponse to be used in models, (2) predictive mean modeling, and (3) final assignment of imputed values using these predictive means. Drug use variables have an additional step to randomly assign the date of first drug use. QC measures were performed at each of these steps. In addition to the checks listed below, all SAS®[144] programs, which were ran by members of the imputation team, were subsequently reviewed by at least two team members for obvious errors. Messages in the SAS® log file, model convergence, and missing values were some of the noticeable errors that were examined. The imputation team also edited demographic, household composition, and income variables. QC measures were implemented in every step of the editing; however, the QC procedures that were used in the editing process will not be discussed in this chapter.[145]

H.2 Step 1. Weight Adjustment for Item Nonresponse to Be Used in Models

In this step, it was necessary to define a set of variables where item nonresponse was characterized. To have been classified as a "complete" respondent, a person would have had to respond to all the questions within the variable set. Only complete respondents were used to build the models in the next step. As a general practice, the weights were adjusted so that the weights for complete respondents represented the entire domain, where "domain" was defined as the population of interest (e.g., lifetime users aged 12 to 17 years old). This was accomplished by using an item response propensity model, a special case of the generalized exponential model (GEM),[146] which is described in greater detail in Appendix B. For this step, QC measures were conducted as follows:

- The output of the response propensity modeling program was checked for singularities. Any singularities that occurred were investigated, and the model was corrected.

- Checks were performed on the output to see whether the GEM model converged. If it did not, the last iteration of the model had a heading titled: "Calculation Of

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

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

[145] See the logical editing procedures used to created these variables in chapters 4, 8, and 9 of this report; for more details on other editing performed on NSDUH data prior to imputation, see Kroutil (2003a, 2003b, 2003c).

[146] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures.

The Betas - Possible Convergence Problem Check Step Adjustments." If this occurred, one or more variables were dropped, which was determined in a number of ways. First, if the coefficient estimate (beta) for a given covariate was equal to 25 or -25, this meant that a stable estimate was not determined for this covariate, and it should have been dropped. Also, optimally each of the covariates in the item response propensity model should have had values distributed across both respondents and nonrespondents. Those variables with a value for "Tot.Nonresp" of 0 did not have this property, and were removed. If the main variable was dropped, its interaction variables were also dropped. For example, if the variable representing age was dropped, then the interaction between age and gender would also have been dropped.

- An indicator was calculated in the response propensity program that measured the maximum adjustment to the weights. In most cases, the adjusted weights resembled the original weights. If the maximum adjustment was too high (usually greater than 3), this was likely due to an overspecified model, where the adjustment was not performing at an optimum level. Large maximum adjustments were investigated and corrected if possible, so that any final adjustment applied was acceptable.

- The number of people identified as item nonrespondents was recorded. This number should have been the same as the number of people who were excluded from the model-building process.

- Using PROC MEANS, the weighted totals for the independent variables in the model were compared both before and after the adjustment. If these weighted totals were equal, the adjustment procedures worked properly.

- The output was checked for missing values, as well as for positive weights for all observations if the predictive mean modeling procedure was used in SUDAAN®.[147]

- Any changes to existing programs were checked by those who ran the programs, as well as other members of the imputation team.

H.3 Step 2. Predictive Mean Modeling

For each question, modeling procedures were used to determine the predicted mean values for each respondent. For example, a model was used to determine the probability of lifetime usage of a given drug based on the responses to the gate question.[148] Although only item respondents contributed to the model, predicted mean values were determined regardless of whether the respondent answered the question or not. These predicted means were calculated based on Poisson regression models, failure time models, binomial and multinomial logistic models, or ordinary weighted least squares regression models with the response variable

[147] SUDAAN® is a registered trademark of RTI International.

[148] The "gate question" was the first question in the module for a given drug, which asked the respondent whether he or she had ever used the drug.

appropriately transformed. The models are discussed in detail in the main body of this report. For this step, the following quality control measures were employed:

- Many of the independent variables were categorical variables and were subsequently converted into a set of indicator variables in an intermediate step. A list of a few observations on the dataset was printed to ensure that all of the indicator variables were created correctly.[149]

- All models were checked for singularities and collinearities. For any singularities that occurred, they were investigated and the model was corrected.

- For Poisson regression models, failure time models, and logistic models, convergence was ensured by checking the output to see if convergence was obtained. For logistic models, the log file was also checked for "data warning" messages or other SUDAAN®-specific errors.[150] If there was a "data warning" message in the log, the SUDAAN® model was unstable and variables were removed to produce stability in the estimates. Similar to the response propensity model, if the main variable was dropped, its interaction variables were also dropped.

- Output was checked to verify that everything worked properly in the regression model.

- If there were two models in the frequency modeling programs, the convergence in both models were checked.

- For age at first use in the drug section, the predicted age at first use was crossed with the respondent's age. The integer portion of the predicted age at first use could not have exceeded the respondent's age. Also, a subset of observations on the output dataset was carefully investigated to ensure that all of the predicted values and indicators were logical.

- A check was made to ensure that each respondent in the domain had a valid predicted mean.

- Any changes to existing programs were checked by those who ran the programs, as well as other members of the imputation team.

H.4 Step 3. Final Assignment of Imputed Values

The predicted means from Step 2 were used to determine the final assignments of imputed values in a hot-deck step. The goal of this step was to make donors and recipients as similar as possible. A neighborhood of potential donors was used, if possible, so that the donor

[149] Although the CLASS statement could have been used in SAS® to automatically create the appropriate indicator variables, no such option was available in SAS®-callable SUDAAN® (Release 8.0), which was used to fit the polytomous logistic regression models.

[150] Greater details can be found in the *SUDAAN User’s Manual: Release 8.0* (RTI, 2001).

selected was different each time the procedure was ran. However, all potential donors in a neighborhood needed to have very similar predicted means. Quality control checks in this step had two objectives: (1) to ensure that the imputed values were consistent with preexisting nonmissing values and (2) to ensure that the imputed values were assigned as intended.

The provisional and final univariate imputations[151] included the following sets of variables: the Hispanic origin indicator, binary and finer income variables, constituent variables method for imputing health insurance variables, questionnaire roster variables, lifetime usage of various drugs, recency and frequency of use of various drugs, and age at first drug use. For these univariate imputations, the output was checked for the items given in the following list. Also, separate QC programs were created for finer income, and age at first drug use univariate imputation programs to check the following output results.

- Unusual imputed values were noted. If the imputed value was equivalent to one of the standard NSDUH missing value codes, this signaled a failure to obtain a donor, and measures were required to revise the programs so that a donor could have been found. If the imputed value was otherwise unusual, the imputation process was examined to ensure that no error occurred.

- The number of cases that had a neighborhood size with a donor within 1 percent was noted.

- The distribution of imputed values by edited values was checked to see if the imputed values were correctly assigned in each imputation class.

- The number of cases that were imputed within various levels of restrictiveness of the likeness constraints (as determined by the variable SMALLFLG) was noted.[152]

- The imputed values were crossed with the imputation indicators to ensure that the indicators were created correctly.

- The frequency of the variable "WORKED" was checked to ensure that no values were equal to zero. Values greater than zero signified that the imputation procedure was able to find a donor for all missing cases.

- The imputed values were checked against preexisting nonmissing values for consistency. Listed below were a few checks carried out to ensure the consistency.

[151] Provisional univariate imputations include binary income, "Constituent Variables Method" for health insurance for various types of health insurance, drug lifetime, drug recency/frequency. Final univariate imputations include Hispanic origin indicator, finer income, "Constituent Variables Method" for health insurance for any other health insurance, roster, drug age at first use.

[152] Refer to Appendix F for more details about likeness restrictions and the "SMALLFLG" variable.

- ☐ The imputation-revised age at first use was crossed with respondent's current age to ensure that the age at first use was never greater than the respondent's age.

- ☐ If there were one or more child[153] drugs, the imputed variables of the parent drug were crossed with those of the child drug(s) to ensure consistency.

- ☐ For parent-child drugs, the parent drug's age at first use must have been less than or equal to the child drug’s age at first use.

- ☐ The respondent's age at first drug use must not have equaled the respondent’s age, if the recency was "not in the past year."

- ☐ The imputed number of people in household under age 18 should have been within a lower and upper bound based on the value of imputed household size and the nonmissing ages in the roster.

- ☐ In binary income variable imputations, donors and recipient were required to have the same value for whether the respondent has family members in household (irfamskp).

- ☐ For income finer category, made sure that finer category was consistent with binary category.

- The edited variables were crossed with imputed variables to ensure that the imputations were carried out correctly. For example, edited number of people in household aged 65 or older (HH65) was compared with imputed number of people in household aged 65 or older (IRHH65) to ensure that IRHH65 had no missing values.

- Any changes to existing programs were checked by those who ran the programs, as well as other members of the imputation team.

Multivariate imputations were performed on the following sets of variables: some of the demographic variables (with multinomial cells), binary income variables, health insurance variables (both the "Old Method" and the "Constituent Variables Method"), lifetime drug use, and recency and frequency of drug use. For these multivariate imputations, the items given in the following list were checked. Also, separate QC programs for drug multivariate imputation program were created to check the output for each drug.

- Any missing values were noted. This occurred when the program was unsuccessful in assigning an imputed value, such as, drug recency (1, 2, 3, 4, 9), 30-day frequency (1-31, 91, 93), or 12-month frequency (1-365, 991, 993).

[153] A parent/child drug relationship occurred in modules that included subgate questions of substances that were of interest in their own right. For example, in the hallucinogens module, there was interest in the usage of LSD, PCP, and Ecstasy, which were all considered as "child" drugs of the "parent" drug hallucinogen.

- Any cases where the imputed value was not consistent with preexisting nonmissing values were noted. Those were cases where one or more variables were imputed, and one or more of these variables violated one or more of the following conditions:

  - ☐ The 12-month frequency must have equaled or exceeded the 30-day frequency.
  - ☐ Past month users must have had a valid 30-day frequency (not a skip code).
  - ☐ Past year users must have had a valid 12-month frequency (not a skip code).
  - ☐ For alcohol, 30-day frequency must have exceeded or equaled the "binge" drinking frequency.
  - ☐ For parent-child drugs (e.g., cocaine and crack, smokeless tobacco, and snuff), the parent drug recency must have occurred no later than the child drug's recency.
  - ☐ For cocaine and crack, the cocaine 12-month frequency must have equaled or exceeded the crack 12-month frequency, if it existed.
  - ☐ For cocaine and crack, the cocaine 30-day frequency must have equaled or exceeded the crack 30-day frequency, if it existed.
  - ☐ The recency and frequency of use variables that were imputed must have been consistent with the time period between the birthday and interview date, as well as the time period between the interview date and the month that the respondent began using, if that variable was available. For example, if the respondent was not a past month user, the imputed 12-month frequency of use could not have exceeded the maximum usage period less 30.
  - ☐ If the respondent's age was equal to the age at first use, the recency of use must have been imputed to be past month or past year not past month.
  - ☐ For past month users, the 30-day frequency must have exceeded the 12-month frequency less 335.
  - ☐ If the edited age at first use was equal to the current age of the respondent, the imputed recency must have been consistent with the time period between the birthday and the interview date, and it must have been consistent with the month that the respondent began using, if available.

- ☐ For income, only people who answered "yes" to either the welfare payments or other welfare services source of income questions had valid answers concerning months on welfare.

- ☐ For health insurance, respondents who indicated that they had health insurance, but were missing the private health insurance indicator required donors who had some health insurance.

- The distribution of the imputed values was compared with the distribution of nonimputed values. Unusual patterns in these distributions were investigated. For example, this included the distribution of lifetime users versus nonlifetime users, the distributions of recency and frequency of use, and the age at first use distributions for drugs. For income, this included the distributions of family income variables.

- Looking at all respondents, the distribution of values was regarded after imputation had been implemented.

- It was necessary to ensure that everyone, to whom the variable did not apply, received a skip code for the final imputed variable. For example, all those in the age group 12–14 should have had a value of 99 for the imputation-revised marital status variable, IRMARIT.

- It was necessary to ensure that any restrictions on the final imputed value for a given nonrespondent were honored. For example, some respondents were known to have been employed, but either full-time or part-time employment status was not known. Checks were carried out to ensure these respondents had either full-time or part-time status assigned to variable EMPSTAT4, but not unemployed or other statuses.

- Each pattern of missingness was treated separately. The distribution of imputed values within each missingness pattern was investigated. For example, if it was known that a respondent was a past year user, both past month and past year users were expected among the imputed values, not just past month users.

- The imputed values were crossed with the imputation indicators to ensure that the indicators were created correctly.

- For the recency and frequency of use, provisional imputed values were used in the process before a final vector of predicted means was created. The provisional imputed recencies were crossed with the edited and final imputed recencies by the imputation indicator. If something went wrong in the final multivariate hot-deck step, this check was established to identify it.

- The distribution of edited variables was compared with the distribution of imputed variables to make sure that each imputed value was within the appropriate range corresponding to the value of the edited variable.

H.5 Additional Step for Drug Variables: Assignment of the Date of First Drug Use

For the age at first drug use imputations, an additional step was required that assigned a date of first use. Quality control checks in this step had two objectives: (1) the assigned date must have been consistent with the imputed age at first use, and (2) the assigned date must have been consistent with other imputation-revised drug variables, such as recency and frequency variables.

- The assigned date of first use should have been consistent with the given birth date and the imputation-revised age at first use.
- The assigned date of first use should have been consistent with the given interview date and the imputation-revised recency/frequency of use variables.
- Respondents failing either of the two preceding checks were carefully examined. Occasionally, the error was unavoidable (e.g., when the age at first use, recency of use, and interview date were inconsistent by only 1 day), even after editing. In particular, this could have occurred if the birthday or interview date occurred on the first of the month. It was important to ensure that all inconsistencies that appeared were of this type.
- The imputation-revised year and month of first use were crossed with the edited year and month of first use to ensure that all valid edited year/months were being carried over to the imputation-revised year/month of first use.
- A frequency of the imputation-revised month/day/year of first use variables was run to ensure that all were within the acceptable numbers (i.e. month was between 1 and 12, or 99 for never used).
- If there were one or more child drugs, the imputed variables of the parent drug were crossed with those of the child drug(s) to ensure the consistency.

Sometimes an error was discovered further along in the process, so that a patch was necessary for earlier imputations. When the variables were reimputed and the dataset was updated, it was crucial to compare the old (incorrect) imputation-revised variable and the new corrected variable with the reimputed values. This was necessary to ensure that (1) the changes made were within expectation, and that (2) other cases did not inadvertently change with the correction. Cases that had unanticipated changes were investigated individually.

In addition, all imputation-revised variables and imputation indicators were checked to ensure that each variable label was correct and the length of the variable was acceptable.

For all of the programs, any changes to existing programs were checked by those who ran the programs, as well as other members of the imputation team.

Appendix I: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

Appendix I: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

I.1 Introduction

In the household roster for the 2002 National Survey on Drug Use and Health (NSDUH),[154] the interviewer was supposed to enter a roster of the respondent's entire household, which included age, gender, and the relationship to the respondent. It was not uncommon for the interviewer to enter a relationship code, age, or gender that did not make sense based on the age and gender of the respondent given in the core part of the questionnaire. Previously in the survey, when the computer-assisted interviewing (CAI) instrument was first implemented, such responses would have been flagged at the data processing stage. Since the age and gender of the respondent given in the core part of the questionnaire were not allowed to change, the relationship code and sometimes the age of the roster member were set to bad data. However, in later survey years and especially in the 2002 survey, consistency checks were added to the instrument that allowed the interviewer, if needed, to correct the error while giving the interview. Details about these consistency checks are presented in Chapter 8 of the main body of this report.

In general, two types of consistency checks were implemented in the 2002 survey. The first type compared the entry in the roster with previously entered questionnaire information, specifically the respondent's age (CURNTAGE) and gender, and the second type checked for internal consistency within the household roster. In some cases, a consistency check would have been triggered even though the response was legitimate. This occurred if CURNTAGE was considered incorrect, or in extremely rare family situations, such as a stepmother who was younger than her stepson. With the exception of the check against the previously entered respondent's gender, the interviewer could have overridden the consistency check and explain why the response given was correct. In some cases, the interviewer was correct in overriding the consistency check. In others, however, it was clear that the interviewer misunderstood how the roster should have been put together, and the override to the consistency check was not legitimate.

This appendix summarizes the explanations given by interviewers for consistency check overrides in the household roster. It is divided into two parts: consistency check overrides involving CURNTAGE, and those involving internal consistency checks.

I.2 Override Comments from Interviewers: Comparisons with CURNTAGE

When an interviewer entered the respondent's roster entry (the "self" entry), if the age did not match the age previously entered in the questionnaire, a consistency check was triggered. The comparison was between the roster age for the "self" and CURNTAGE, the value of age that

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

was stored by Blaise.[155] Explanations given by interviewers for overrides to consistency checks against CURNTAGE are provided in Exhibit I.1. Since CURNTAGE had the potential to change constantly throughout the questionnaire, no final variable with this name was created. However, in most cases, the value of CURNTAGE when the roster commenced was equivalent to NEWAGE, the value of CURNTAGE after the drug modules had been completed. In theory, NEWAGE was not always equivalent to the final questionnaire-edited age (AGE), the derivation of which is described in Chapter 4 of the main body of this report. Among the cases listed in Exhibit I.1, however, NEWAGE, AGE, and (presumably) CURNTAGE at the time the roster commenced, are all equivalent.

The explanations given in Exhibit I.1 were not reviewed when determining AGE, nor were they reviewed when determining the final value for the age of the "self" entry in the roster. Any override to a consistency check was ignored; the final age corresponding to the "self" entry in the roster was set to AGE, which, along with the screener age, is provided in this exhibit. Hence, the "original roster age for self" was in fact confirmed in a consistency check, but was still ignored. The "Comments" column in Exhibit I.1 indicates whether the explanation disavows the override of the consistency check and supports the value of CURNTAGE for the respondent's age, or whether the override was correct and CURNTAGE was wrong. The last column in Exhibit I.1 indicates whether the roster of the other pair member, if it exists, supports CURNTAGE or the override age as the respondent's age. In this exhibit, the comparison is made with AGE rather than CURNTAGE, since they are equivalent. In almost all cases, the difference between CURNTAGE and the override age is 1 year or less. Differences between a "correct" age and the final value for AGE of 1 year or less are not of major concern. However, in a few cases, the difference between CURNTAGE and the override age exceeds 1 year. Therefore, the correctness of the final value for AGE is questionable for these cases. The amount of discrepancy is in bold text in the Comments column.

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

Exhibit I.1 Explanations for Overrides to Consistency Checks against CURNTAGE

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE= Final Roster age	Verbatim Explanation from Field Interviewers[1]	Comment	Roster of Other Pair Member Supports
1	37	38	38	37	used wrong age on newton	Indicates override should be ignored	Not in a pair
2	39	38	38	39	he didn.t remember his exact age	One is wrong, not clear which	AGE
3	39	38	38	39	39 is correct current age	Indicates override should be ignored	AGE
4	20	19	20	20	r is 20, misunderstood "age on your last bd"	Indicates override should be ignored	AGE
5	13	12	13	13	the child is providing conflicting statistic on his age	One is wrong, not clear which	Not in a pair
6	12	13	13	12	R is 13	Indicates AGE is wrong by 1 year	Not in a pair
7	24	25	25	24	24	Indicates override should be ignored	Neither
8	22	23	23	22	the ancle didn't know for sure how old his nice was	One is wrong, not clear which	Not in a pair
9	32	33	33	32	householder was not sure what age she was	One is wrong, not clear which	Override
10	18	19	19	18	true age is 19	Indicates AGE is wrong by 1 year	Override
11	24	16	16	24	gave	Problem: age difference is large	Override
12	25	23	23	25	sc reening respondent did not kn ow correct ages	Indicates override should be ignored	Override

(continued)

Exhibit I.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE= Final Roster age	Verbataim Explanation from Field Interviewers[1]	Comment	Roster of Other Pair Member Supports
13	23	22	22	23	r is 23	Indicates override should be ignored	Override
14	22	26	22	22	22	Indicates override should be ignored	Not in a pair
15	25	27	25	25	hh had given wrong relationship	Is the roommate the true respondent?	Not in a pair
16	21	22	22	21	res is 22 years old	Indicates AGE is wrong by 1 year	Override
17	24	23	24	24	wrong answer was given before. r is 23.	Indicates AGE is wrong by 1 year	Override
18	35	34	34	35	he is actually 34	Indicates AGE is wrong by 1 year	Override
19	13	12	12	13	respondent was earlier stated as 12	Indicates AGE is wrong by 1 year	Override
20	44	43	43	44	r is a male		Override
21	75	74	75	75	the date of birth is correct, however, the computer kept saying she was 75, even when we wen	Indicates override should be ignored	AGE
22	22	21	21	22	she does not want to answer		Not in a pair
23	43	42	42	43	she is 42	Indicates AGE is wrong by 1 year	Override
24	19	18	19	19	age of 18 is correct	Indicates AGE is wrong by 1 year	AGE
25	46	47	47	46	wrong age	One is wrong, not clear which	AGE

(continued)

Exhibit I.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE= Final Roster age	Verbataim Explanation from Field Interviewers[1]	Comment	Roster of Other Pair Member Supports
26	33	32	32	33	age is 32	Indicates AGE is wrong by 1 year	Not in a pair
27	47	44	44	47	r gave incorrect age earlier	Indicates AGE is wrong by **3 YEARS**	Override
28	27	28	28	27	respondent is 28	Indicates AGE is wrong by 1 year	Override
29	41	42	41	41	at begin r said 41, correct age 42	Indicates AGE is wrong by 1 year	Override
30	28	27	27	28	Birth date was May 5, 1974 making the R 28 years old.	Indicates override should be ignored	AGE
31	20	21	21	20	Respondent is 21 years old.	Indicates AGE is wrong by 1 year	Not in a pair
32	47	48	48	47	48 was correct lady confused the subject	Indicates AGE is wrong by 1 year	AGE
33	15	14	15	15	interviewer error- oops! R is 15	Indicates override should be ignored	Not in a pair
34	21	22	22	21	rsp didn t know exact	One is wrong, not clear which	Override
35	32	33	34	32	says he is 33 now and was born in 1969	Indicates AGE is wrong by 1 year	Not in a pair
36	18	22	18	18	answer is self instead of unmarried partner	Indicates AGE is wrong by **4 YEARS**	Override
37	40	24	24	24	R is 24 years old	Matches with AGE; bad NEWAGE	Not in a pair

(continued)

Exhibit I.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW-AGE	Original Roster Age for Self	Screener Age	AGE= Final Roster age	Verbataim Explanation from Field Interviewers[1]	Comment	Roster of Other Pair Member Supports
38	21	23	23	21	FEMALE IS 23	Indicates AGE is wrong by **2 YEARS**	Override
39	33	35	33	33	the resp 'sex is correct		AGE
40	33	23	23	33	notice the age are correct		Not in a pair
41	22	21	21	22	Age is 22	Indicates override should be ignored	AGE
42	28	29	29	28	R,WHEN SCREENING SAID SISTER 29 YEARS.R,ONLY 28 NOW WILL BE 29 BIRTHDAY ALSO SISTER 28,27 26	Indicates override should be ignored	Override
43	24	22	22	24	r is clearly confused with year of birth	One is wrong, not clear which	AGE
44	38	37	38	38	R's incorrect answer	One is wrong, not clear which	AGE

[1] These entries came directly from the 2002 NSDUH Field Interviewers. Any typographical errors or misspellings were transcribed directly and not corrected.

I.3 Override Comments from Interviewers: Internal Consistency Check Overrides

New consistency checks that were added for the 2002 survey also checked for internal consistency in the roster. Explanations by interviewers for overrides to internal consistency checks are given in Exhibit I.2. These explanations were not ignored, but rather were individually evaluated to determine their legitimacy. Also provided in this exhibit are the questionnaire-edited age of the respondent (AGE), the age and relationship to the respondent of the roster member in question, and, in the "Comment" column, an evaluation of whether the override was considered legitimate. If the override was legitimate, no edit was applied to the age or relationship code of the roster member. On the other hand, if the override was not considered legitimate, the override was overruled and the relationship code, and sometimes the roster member's age, was set to bad data. In this instance, a brief indication of the probable true relationship of the roster member to the respondent is given in the "Comment" column of the table.

Exhibit I.2 Explanations for Overrides to Internal Consistency Checks

#	Consistency check	AGE	Roster member's age and relationship to respondent	Verbataim Explanation from Field Interviewers[1]	Comment
1	Respondent's son is older than respondent	22	24-year-old stepson of respondent	22 yr old is married to 43 yr old woman who has these sons that are older	Legitimate; interviewer's override stands
2	Respondent's mother is younger than respondent	43	33-year-old stepmother of respondent	this is the r. mother through father second mirrage	Legitimate; interviewer's override stands
3	Respondent's father is younger than respondent	31	13-year-old stepfather of respondent	unknown	Overrule; probable stepson of respondent
4	Respondent's son is older than respondent	12	14-year-old stepson of respondent	step son is not biological	Overrule; probable stepson of respondent's parent
5	Respondent's father is younger than respondent	42	11-year-old adoptive father of respondent	r is the adoptive father	Overrule; probable adoptive son of respondent
6	Respondent's grandfather is younger than respondent	16	6-year-old grandfather of respondent	correc ted	Overrule; either the age or relationship code is wrong; other pair member indicates error is in relationship code

[1] These entries came directly from the 2002 NSDUH Field Interviewers. Any typographical errors or misspellings were transcribed directly and not corrected.