1999 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

IMPUTATION REPORT

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

Authors:

Eric A. Grau
Katherine R. Bowman
Katherine E. D. Giacoletti
Dawn M. Odom
Neeraja S. Sathe

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute
RTP, North Carolina 27709

July 13, 2001

1999 NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE

IMPUTATION REPORT

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

Authors: Project Director: Thomas G. Virag

Eric A. Grau
Katherine R. Bowman
Katherine E. D. Giacoletti
Dawn M. Odom
Neeraja S. Sathe

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute
RTP, North Carolina 27709

July 13, 2001

Acknowledgments

This report would not be possible without the guidance and input of staff from the Office of Applied Studies (OAS). In particular, useful comments were provided by Dicy Butler, Joe Gfroerer, and Art Hughes. Special thanks are also due to several Research Triangle Institute (RTI) staff members. Avinash Singh and Ralph Folsom, along with Eric Grau, codeveloped the predictive mean neighborhood (PMN) methodology. Dr. Folsom and colleagues provided the text for Appendix B, and Dr. Singh and colleagues provided the text for Appendix C. Larry Myers was instrumental in the implementation of the failure time models for the finer category income imputations and provided most of the content for Sections 7.2.4.3 and 7.2.4.4. Peter Frechtel, Amy Licata, and Seungho Huh provided many helpful comments and did considerable work on Appendices I and J. Finally, Richard Straw copyedited and formatted the report in preparation for publication, and Brenda Porter helped prepare some of the tables in the appendices.

Table of Contents

Chapter **Page**

Table of Contents (Continued)

Chapter **Page**

Table of Contents (Continued)

Chapter **Page**

Table of Contents (Continued)

Chapter **Page**

Table of Contents (Continued)

Table of Contents (Continued)

List of Tables

Table **Page**

List of Exhibits

Exhibit **Page**

This page intentionally left blank.

1. Introduction

In 1999, the National Household Survey on Drug Abuse (NHSDA) was implemented using a new 50-State design. Other major changes to the study protocol included the introduction of computer-assisted interviewing (CAI) methods for both screening households and interviewing selected respondents. Prior to 1999, paper-and-pencil interviewing (PAPI) was used both for screening and interviewing.

The 50-State design was developed to allow the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide direct estimates for eight large States and estimates based on small area estimation methods for the remaining States and the District of Columbia (DC). This resulted in a major increase in sample size at the national level (from about 20,000 per year to 70,000 per year). In addition, a national sample was surveyed using the PAPI methodology during the same time period.

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

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

Although the 1999 NHSDA contained both a CAI sample and a PAPI sample, the focus of this report is on the procedures implemented on the CAI sample only. The eligibility and completeness criteria for both CAI and PAPI are discussed in Chapter 2, followed by a summary of the imputation procedures implemented for both samples in Chapter 3. The eligibility and completeness criteria for PAPI, as well as the imputation methodology, did not differ substantially from previous years. Details of imputation procedures applied on the PAPI sample can be obtained in the *1998 NHSDA Methodological Resource Book*, Vol. I, Section 3 (OAS, 1999). Chapters 4 and 5 detail the imputation procedures applied to the core and noncore demographic variables, respectively. The drug imputation procedures are discussed in Chapter 6. Most of the editing procedures that were applied to the demographic and drug variables discussed in Chapters 4, 5, and 6 are summarized by Kroutil (2001a, 2001b, 2001c). The editing procedures for the income and household composition variables, however, are discussed in this document. Chapter 7 summarizes the editing and imputation procedures applied to the income variables, and Chapter 8 details the edits applied to the household roster, as well as the creation and imputation of missing values in the roster-derived household composition variables.

This document also contains 11 appendices, including 4 summaries of the various imputation methodologies that have been used in past years and in the current CAI sample. The hot deck is described in **Appendix A**; the model-based imputation method used in PAPI and in previous years is summarized in **Appendix B**; the general model used to adjust weights for item nonresponse is discussed in **Appendix C**; and the new methodology implemented in the 1999 CAI sample, the predictive mean neighborhood (PMN), is described in **Appendix D**. Respondents had the opportunity to write in responses to some of the drug and demographic questions if they felt the given responses did not apply. These responses, called "alpha-specify" responses, were coded so that the data could be summarized in a meaningful way. A discussion of how this was done for race-Hispanicity and employment status is described in **Appendices E** and **F**. (Coding of alpha-specify responses for other variables is summarized by Kroutil, 2001a, 2001b, 2001c.) A summary of the models used in the PMN methodology for various variables is given in **Appendix G**.

An error in the first run of the imputation procedures, as applied to drugs, was detected. A description of the error and how it was corrected is given in **Appendix H**. This is followed by a summary of the number of respondents who met various constraints that could be loosened in the imputation process in **Appendix I**. **Appendix J** gives details of the vector of predicted means used in the multivariate PMN procedure for drugs and health insurance for various patterns of missing values, in addition to the logical constraints required, and the CAI questionnaire is given in **Appendix K**.

This page intentionally left blank.

2. Eligibility and Completeness Rules

2.1 Eligibility Criteria

The population of eligible respondents for the 1999 NHSDA was all civilian, noninstitutionalized residents of the United States (including DC) aged 12 or older. As in other recent NHSDAs, this population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Persons excluded from the 1999 survey included those with no fixed household address (e.g., homeless transients *not* in shelters), residents of institutional group quarters, such as jails and hospitals, and active military personnel.

During screening, respondents were asked to identify all eligible household members so that only eligible individuals were listed and therefore potentially selected. Due to screening errors, some ineligible individuals were selected, however, and later determined to be ineligible at the time of interview. For a summary of the number of eligible persons rostered and completed interviews obtained in the 1999 NHSDA for both the CAI and PAPI samples, see **Table 1**.

Table 1. Household and Person Eligibility and Response Rates, 1999 NHSDA

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases
CAI	223,868	187,842	169,166	351,396	89,883	67,096	66,706
PAPI	46,328	40,584	35,635	75,084	18,896	13,886	13,809

2.2 CAI Completed Case Rule

To be considered a completed case for purposes of analysis, a CAI respondent had to provide "yes" or "no" answers to the cigarette gate question and at least 9 of the other 14 gate questions. Unlike the PAPI questionnaire in 1999 and prior NHSDAs, no logical inference could be made from information within a section if the gate question was not answered because the CAI instrument routed respondents out of a section if the gate question was not answered. (The PAPI completed case rule, as in previous years, required responses to the alcohol, marijuana, and cocaine gate questions.) For a summary of the number of completed cases in the 1999 NHSDA for both CAI and PAPI, see Table 1.

This page intentionally left blank.

3. Overview of Item Imputation Procedures

3.1 Introduction

As in most large-scale sample surveys, the 1999 NHSDA faced the problem of analyzing datasets that contained missing responses for some items. Associated with this problem of item nonresponse is that of inconsistent or invalid responses and violation of skip patterns. Although the CAI instrument enforced skip patterns (which would be expected to reduce inconsistencies) and performed some consistency checks, inconsistent and invalid responses still occurred. These response errors are an obvious source of bias that must be considered in data analysis (Cox & Cohen, 1985, p. 238).

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, pp. 224-225). For instance, at times gender can be inferred from the 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.

To maintain consistency with the practice used in previous NHSDAs, the procedures implemented in the 1999 PAPI supplemental sample were identical to those from previous years. However, in the spirit of the effort to improve the quality of estimates from the redesigned NHSDA, and as a result of fundamental differences between PAPI and CAI, there was a need to change the way missing data were edited and imputed. Thus, new imputation procedures were developed and implemented for the CAI sample. Aside from changes in the treatment of core demographics (notably race/Hispanicity and employment), most of the changes affected the imputation of missing values in the drug, income, insurance, and household composition (roster-derived) variables. **Exhibit 1** provides a brief summary of the type of imputation procedure used for each of the variables imputed in the 1998 and 1999 NHSDAs.

3.2 Overview of Item Imputation Procedures in the PAPI Supplemental Sample

The unweighted sequential hot-deck method of statistical item imputation was used for the PAPI supplemental sample of the 1999 NHSDA for items that exhibited low item

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

Variable	1998 PAPI	1999 PAPI	1999 CAI
All Rostered People from Screener			
Age	HD[1]	HD	×[2]
Race	HD	HD	×
Gender	HD	HD	×
Hispanic-Origin Indicator	HD	HD	×
Marital Status[3]	HD	n/a[4]	n/a
Completed Cases Only			
Race	n/a	n/a	HD
Hispanic-Origin Indicator	n/a	n/a	HD
Hispanic-Origin Group	HD	HD	HD
Marital Status	n/a	n/a	HD
Work Status (12 levels)	HD	HD	n/a
Employment Status (5 levels)	Derived[5]	Derived	HD
High School Graduate Indicator	HD	HD	n/i[6]
Education	HD	HD	HD
Private and Total Health Insurance	HD	HD	MPMN[7]
Drug Lifetime Usage (enters into recency)			UPMN[8]
Drug Recency of Use	HD	HD	MPMN
ALC, MRJ, COC Frequency-of-Use (12 months)	Model[9]	Model	MPMN
Other Drug Frequency-of-Use (12 months)	HD	HD	MPMN
Drug Frequency-of-Use (30 days)	n/i	n/i	MPMN
Binge Drinking Frequency (30 days)	n/i	n/i	MPMN
Age at First Use	n/i	n/i	UPMN
Age at First Daily Cigarette Use	n/i	n/i	UPMN
Personal and Family Income Binary Variables	Model	Model	MPMN
Personal and Family Income Finer Categories	Model	Model	UPMN
Household Size (Roster-Derived Variable)	n/i	n/i	UPMN
Other Household Composition (Roster-Derived) Variables	HD	HD	UPMN

[1] "HD" refers to the unweighted sequential hot-deck method of item imputation.
[2] The "×" represents no single imputation for all rostered people from the screener. In the CAI sample, missing values for age and gender among completed cases were not replaced by imputation. Imputations of the race and Hispanic indicator variables were performed separately for all rostered persons and for completed cases (see **Chapter 4**).
[3] In 1999, marital status was removed from the screener, so that it was only necessary to impute marital status missing values for completed cases.
[4] "n/a" refers to data unavailable for the NHSDA study or imputed for a different set of respondents.
[5] "Derived" reflects the fact that the five-level employment status variable was a recode of other imputed variables in previous NHSDAs (see **Chapter 5**).
[6] "n/i" refers to data that were available but not imputed.
[7] "MPMN" refers to a new "multivariate predictive mean neighborhood" model-based procedure described in this report.
[8] "UPMN" refers to a new "univariate predictive mean neighborhood" model-based procedure described in this report.
[9] "Model" refers to model-based item imputation, as applied to the 12-month frequency of use variable in previous NHSDAs.

nonresponse (i.e., the demographic variables, drug recency-of-use variables, and some of the drug frequency-of-use variables). Hot-deck imputation involves the replacement of a missing value with a valid code taken from another respondent who is "similar" and has complete data. Responding and nonresponding cases were sorted together by a variable or collection of variables closely related to the variable of interest (Y) that are known for both item respondents and item nonrespondents. When the donor and recipient are required to have the same values of a sorting variable, the variable is called a "classing variable." For sequential hot-deck imputation (Little & Rubin, 1987, p. 62), a missing value of Y is replaced by the nearest responding value preceding it in the sequence. For example, recency of alcohol, marijuana, and cocaine use are good candidates for classing or sorting variables when imputing other drug recency and frequency-of-use items because alcohol, marijuana, and cocaine recency will be known for all interview respondents.[1] A complete discussion of the hot-deck method of imputation is presented in **Appendix A**.

As in the 1992 to 1998 NHSDAs, a model-based approach was used for imputing missing data for items in the PAPI supplemental sample that exhibited large nonresponse or that could potentially cause large amounts of nonresponse bias (e.g., the alcohol, marijuana, and cocaine 12-month frequency-of-use questions). For any nonrespondent, this approach will produce an estimate of the probability that the nonrespondent would have given each of the possible answer categories. These nonrespondent answer category probabilities were estimated using a polytomous logistic regression model.

With the model-based approach, the probabilities from the polytomous model are used to randomly generate a definitive response category for each nonrespondent. The major analysis variables (age, race, ethnicity, and gender), as well as the recency variables for alcohol, marijuana, and cocaine, are good candidates for explanatory variables in these models. A more complete discussion of the model-based method of imputing categorical responses for the PAPI supplemental sample is provided in **Appendix B**.

3.3 Overview of Item Imputation Procedures in the CAI Sample

As stated in the introduction to this chapter, many changes were instituted in the treatment of missing and inconsistent values in the CAI sample compared with previous NHSDAs. Imputation procedures for basic demographics did not change appreciably although some minor changes were implemented prior to the imputation of missing values. The biggest

[1]By definition of PAPI, "respondent" or "completed case."

changes in the imputation of missing values for basic demographic variables occurred with the handling of race/Hispanicity and employment. More scrutiny was given to "alpha-specify" responses to the race and Hispanicity questions (written responses entered by the interviewer in cases where the given numeric responses were seen not to apply). Major changes were required in the handling of the employment variables because the structure of the questionnaire dramatically changed from PAPI to CAI. However, missing values for all demographics except the income and insurance variables were imputed using an unweighted sequential hot-deck procedure, as in past years.

The most significant changes in the imputation procedures were evident with the core drug, income, insurance, and household composition (roster-derived) variables. In previous years, recency of use and 12-month frequency of use were the only measures with missing values replaced by imputed values. However, with the implementation of the new procedures in 1999, 30-day frequency of use, age at first use, age at first daily use of cigarettes, and 30-day binge drinking frequency were added to the list of variables requiring imputation. Whereas the unweighted sequential hot deck was previously the tool of choice, this procedure was (mostly) abandoned in 1999 in favor of a combination of weighted regression imputation and a nearest neighbor hot-deck imputation (explained in detail in **Appendix A**). The hot-deck donor is determined using the predicted mean from the model, where the hot deck is random whenever possible.[2] This new procedure will henceforth be referred to as predictive mean neighborhood (PMN).

Depending on the response variable, the weighted regression model in PMN was either a binary or multinomial logistic model, or a multiple linear regression with the response variable appropriately transformed. The weights in the model were sample design weights,[3] 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

---

[2]The lack of randomness associated with the unweighted sequential hot deck is alleviated with the random nearest neighbor hot deck (Little & Rubin, 1987, p. 65), where a set of complete records that are "close" to a given incomplete record are candidates to donate values to the incomplete record. The set, or neighborhood, of complete records is defined by the covariates under a suitable metric, then imputed values are chosen at random from the neighborhood of complete records.

[3]In the CAI sample of the 1999 NHSDA, the final analysis weights were used if they were available. However, because the modeling of the final nonresponse adjustment was not completed at the time of the drug imputations, the person-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.

Appendix C. Predicted values (predictive means) were obtained from the model for both item respondents and item nonrespondents. The predictive means were used to determine the neighborhoods for the nearest neighbor hot-deck procedure, from which donors were randomly selected for the final assignment of imputed values. This assignment was either done one value at a time (a univariate predictive mean neighborhood [UPMN]) or using several response variables at once (a multivariate predictive mean neighborhood [MPMN]). Details of these imputation procedures are given in **Appendix D**.

The justification for changing the procedures followed from a number of general shortcomings associated with unweighted sequential hot deck as it was applied in previous NHSDAs (and the 1999 PAPI supplemental sample):

1. The first few sorting covariates (often only the first and second covariate) almost entirely determine what donor will be used for a particular respondent with missing data, regardless of how many sorting covariates are included.

2. Information about the variable of interest (*y*) is not used in defining the potential donor (the neighborhoods obtained using the PMN techniques are defined based on predictive means, which are in turn determined using *y*).

3. The choice of donor, after the sort has been completed, is deterministic, making it difficult to determine the variance due to imputation.

4. In most cases, weights are not used to determine the most appropriate donor for a respondent with missing data.

5. No attempt is made to ensure that imputed responses are consistent with nonimputed responses for a given case.

6. The correlation among responses across several variables is mostly lost with this method (with PMN, correlations among covariates are maintained in the model coefficients). Correlations across response variables are maintained when MPMN is used).

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

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

Although statistical imputation of the drug use variables could not proceed separately within each State due to insufficient pools of donors, information about the State of residence of each respondent was incorporated in the modeling and hot-deck steps of the PMN procedure in the CAI sample. Respondents were separated into three State usage-level categories for each drug depending on the response variable of interest. Respondents from States with high usage of a given drug were placed in one category, respondents from medium usage States into another, and the remainder into a third category. This categorical "State rank" variable was used as one set of covariates in the imputation models. In addition, as another likeness constraint, eligible donors for each item nonrespondent were restricted to be from States with the same level of usage (the same State rank) as the item nonrespondent. A State rank variable was used in a similar manner in the income imputations, both in the modeling and hot-deck steps. The three State rank categories were defined in terms of the income level of the States: high-income States, middle-income States, and low-income States.

4. Core Demographics

4.1 Introduction

As in previous NHSDAs, several demographic characteristics were needed for all respondents in the 1999 NHSDA. In the CAI sample, core demographic data, which were collected on both the screener and the questionnaire, 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 separately).[4] Thus, prior to any processing of the CAI data, completed cases were identified, and only those cases were included in the subsequent editing, imputation, and analysis of questionnaire data.

The core demographics in the CAI sample of the 1999 NHSDA were age, race, Hispanic-origin indicator, Hispanic-origin group, gender, marital status, and educational level (highest grade completed). This list does not include the high school graduate indicator, which was also part of the core demographics in previous years (and the PAPI supplemental sample in 1999). The only noncore demographic variable imputed for CAI was employment status, the questions for which had changed substantially from previous years (and in the 1999 PAPI).[5]

Missing values for all demographic variables were imputed using an unweighted sequential hot-deck procedure. Prior to imputation, logical editing was performed on some of these variables, filling in some missing values, and reducing the amount of statistical imputation required.[6] Logical editing of variables in the CAI sample was done using only the "other" questionnaire responses, and no noncore information was used to edit core variables.

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

---

[4]See the *1999 NHSDA: Sample Design Report* for a description of the imputation procedures used for screener demographics for the set of all eligible rostered CAI individuals (Bowman, Penne, Chromy, & Odom, 2001).

[5]Income, health insurance, and the household composition (roster-dreived) variables were also statistically imputed, but those variables are discussed in Chapters 7 and 8 rather than with the main demographic variables.

[6]Logical editing undertaken to create base variables for imputation is described in this report; for more details on other editing performed on NHSDA data prior to imputation, see Kroutil (2001a, 2001b, 2001c).

4.2 Geographic and Other Commonly Used Sorting and Classing Variables

4.2.1 Household Type

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

4.2.2 Field Interviewer Region

Field interviewer regions ("FI regions") were created within each State as part of the 1999 NHSDA sample design, and the variable FIREGION identifies each such region within a given State. For more detail regarding FI regions, see the *1999 NHSDA: Sample Design Report* (Bowman et al., 2001).

4.2.3 Segment ID

FI regions were partitioned into clusters of adjacent blocks called "segments." The variable SEGID (Segment ID number) is a two-letter State abbreviation followed by a two-digit FI region and two-digit segment identifier, unique within a given State by FI region combination. SEGID was used in the demographic imputations as a sorting variable. For more information regarding segments, see Bowman et al. (2001).

4.2.4 Household ID

A unique household-level identification number was used as a sorting variable in the demographic imputations.

4.2.5 Design Strata

Six design strata in the 1999 NHSDA were defined at the segment level: high Hispanic (≥50%), high black (≥50%), high minority (≥50%), high white (≥90%), medium white (<90% and ≥75%), and the remainder. These strata were not explicitly used in the design of the

CAI sample,[7] but the classification was created for all segments, both PAPI and CAI. Design stratum was used as a sorting variable for demographic imputations. For more information regarding design strata, see Bowman et al. (2001).

4.2.6 State

States are identified by numeric codes, and this information was used as a sorting variable for some of the demographic imputations.

4.2 7 Population Density (Two Versions)

Two population density variables, PDEN and PDEN2, were used to categorize segments according to 1990 census data, and PDEN2 was used as a sorting variable in some demographic imputations. PDEN2 had five levels: segment in 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. PDEN had three levels: segment in MSA with 1 million or more persons, segment in MSA with fewer than 1 million persons, and segment not in MSA.

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

In the CAI sample, the date of the interview, age, and birth date were required for all completed cases. Some editing of these date values was required to resolve inconsistencies and fill in missing data.

4.3.1 Edited Interview Date (INTDATE)

There were no missing interview dates in the CAI data, but a small number of reported interview dates fell outside the quarter in which a case was fielded or had a reported interview year other than 1999. If an interview date was within 2 weeks of the correct quarter for a given case, the interview date was left as reported to preserve data from questions in latter sections of the questionnaire that refer to time periods in relation to the interview date (e.g., "past 30 days"). These reported dates were considered to be close enough to the correct quarter to have been accurately reported although they indicate that the FI did not follow proper survey

---

[7]Before selecting the sample, implicit stratification was achieved by sorting the first-stage sampling units by an MSA/SES (metropolitan statistical area/socioeconomic status) indicator and by the percentage of the population that was non-Hispanic and white.

procedures. On the other hand, any interview dates further outside the assigned quarter than 2 weeks were considered incorrect and replaced with randomly generated dates within the correct quarter. In these latter cases, all questionnaire data from later sections referencing a time period dependent on the interview date (e.g., past month or past year) were set to missing and later imputed. For a summary of the editing of CAI interview dates, see **Table 2**.

Table 2. Interview Date Editing Summary

From Questionnaire		Assigned Within Quarter	
Frequency	Percent	Frequency	Percent
66,556	99.78	150	0.22

4.3.2 Age

4.3.2.1 Final Edited Continuous Age (AGE). With the new CAI instrument, it was expected that age data would not have to be edited or imputed for any completed cases because the computer was programmed to require either age or date of birth before allowing a selected individual to proceed with the questionnaire. However, a small number of completed cases had reported ages that were inconsistent with other age information (birth date, screener age, questionnaire roster age) or had no reported age and an invalid (but not missing) birth date. The final age variable for CAI (AGE) was created by comparing all available age data (core questionnaire age, reported birth date, questionnaire roster age, and screener age) for a given case and assigning age as follows. AGE =

> reported core questionnaire age (NEWAGE), if nonmissing and within 1 year of age calculated from reported birth date and edited interview date (INTDATE), or if birth date is missing or invalid, the core questionnaire age (NEWAGE) is within 1 year of the questionnaire roster age, else
>
> age calculated from reported birth date, if this differs from NEWAGE by a year or more and is within 1 year of the questionnaire roster age and/or screener age, else
>
> questionnaire roster age, if nonmissing and the reported core questionnaire age (NEWAGE) is missing and the birth date is missing or invalid, else

screener age, if the reported core questionnaire age (NEWAGE) and the questionnaire roster age are missing and the birth date is missing or invalid.

For a summary of the editing to create AGE for the CAI sample of the 1999 NHSDA, see **Table 3**.

Table 3. Age Editing and Imputation Summary

From Core Questionnaire		Calculated from Birth Date		From Questionnaire Roster		From Screener	
Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
66,690	99.98	2	0.00	2	0.00	12	0.02

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

4.3.3 Edited Birth Date (BRTHDATE)

CAI respondents were required to provide their date of birth and/or current age at the beginning of the interview in order to continue. Thus, although a number of cases had missing birth dates, each complete case respondent possessed a current age. The interview date and age variables (for which other information was available in the questionnaire and/or screener data) were created first, then the birth date was calculated from these where inconsistencies existed (either in the raw data or as a result of editing age and/or interview date).

In cases with missing birth dates or for which the birth date was inconsistent with the edited interview date and age, the birth date was calculated based on these edited variables as follows:

1. The integer value from the final edited age was converted to a SAS date value[8] by first adding a fraction of a year (in the form of a randomly generated fraction) and then multiplying by 365.25:

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

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

   BRTHDATE = edited interview date - intermediate age.

See **Table 4** for a summary of the birth date editing.

Table 4. Birth Date Editing Summary

From Questionnaire		Calculated from Final Edited Age	
Frequency	Percent	Frequency	Percent
66,340	99.45	366	0.55

4.4 Demographics Requiring Imputation

In the CAI sample, missing values for the demographics of completed cases were imputed separately from those of all eligible rostered individuals. Moreover, no screener information was used to edit questionnaire demographics for CAI for the completed cases, except in some extraordinary circumstances, which are explained below.

4.4.1 Gender

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

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

reported in the questionnaire, EDSEX was set equal to the gender reported during the household screening.[9] In the CAI sample of the 1999 NHSDA, there were no additional missing values for EDSEX; therefore, no statistical imputation was required. For a summary of item nonresponse and editing for gender, see **Table 5**.

Table 5. Gender Editing and Imputation Summary

From Questionnaire		**From Screener Gender**		**Statistically Imputed**	
Frequency	Percent	Frequency	Percent	Frequency	Percent
66,703	100.00	3	0.00	0	0.00

4.4.1.2 Imputation-Revised Gender (IRSEX). The final version of the gender variable for CAI was called IRSEX. In 1999, no statistical imputation was required to create this variable because gender was determined from questionnaire or screening responses for all respondents. For a summary of item nonresponse for gender, see **Table 5**.

4.4.2 Race

4.4.2.1 Edited Race (EDRACE). Unlike in previous NHSDAs and in PAPI, two core questions in the CAI questionnaire focused on the respondent's race. The first question (QD05) allowed the respondent to select multiple race categories, and the second (QD06) asked the respondent to choose from among those selected in QD05 (if more than one race was selected) the single race that best describes him or her. QD05 has an "other" category, and if this category was the only one chosen in QD05, or was chosen in QD06, the respondent was asked to specify a race. There were 13 answer categories in the CAI race questions; however, the final CAI race variable was collapsed to the same four levels as that of past NHDSAs (and the PAPI sample): American Indian or Alaska Native, Asian or Pacific Islander, black, and white.[10] Also,

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

[10]To collapse the CAI 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.

as in previous years, the final imputed race variable was limited to one category per respondent. However, two additional variables were created for CAI that reflected the changes in the race question (see **Section 4.4.4**).

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

> the single race identified in QD05, if that single race was not "other," else
>
> race recode from alpha-specify response(s)[11] when "other" was the only race selected in QD05, if a valid recode was available,[12] else
>
> missing.

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

> the race response in QD06, if it is not "other" or missing, else
>
> race recode from alpha-specify response if QD06 = "other" and a valid recode is available, else
>
> race assigned from the multiple responses given to QD05, using the following priority: black/African American, Asian, American Indian/Alaskan Native, white.[13]

If no response was given to QD05, EDRACE =

---

[11]Both QD04 (Hispanic-origin group question, see **Section 4.4.6**) and QD05/QD06 allowed interviewers to enter a written response to the questions about the respondent's Hispanic group or race, respectively, when the listed responses were seen not to apply and the category "other" was selected. These written responses are called "alpha-specify" responses, which were coded using the dictionary given in **Appendix E**. 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 E.

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

[13] In order to select one racial group from multiple selected groups, a priority rule was established whereby if Black/African American was among the groups selected then the single race for the respondent is Black/African American, otherwise if Asian was among the groups selected then the single race for the respondent is Asian, etc.

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

missing.

4.4.2.2 Imputation-Revised Race (IRRACE). An imputation-revised race variable was created using an unweighted sequential hot-deck method to impute missing values. Imputation classes were created based on the Hispanic-origin status and (if applicable) Hispanic-origin group of respondents with missing values of EDRACE. The imputation classes were Mexican, Cuban, Puerto Rican, Central or South American, and Hispanic (specific group unknown). For Hispanic respondents with missing EDRACE values, donors were restricted to respondents in the same Hispanic-origin group. For non-Hispanic respondents, or respondents whose Hispanic-origin status was unknown (because missing values of the Hispanic-Origin indicator and group were imputed after the imputation of the race variable), all respondents with valid EDRACE values were eligible to be donors.

The file was serpentine sorted within imputation classes by design stratum, census region, segment, household type, and a random number prior to imputation. The final imputation-revised race variable was called IRRACE. See **Table 6** for a summary of nonresponse for race.

Table 6. Race Editing and Imputation Summary

From Questionnaire		**From Alpha-Specify Codes**		**Statistically Imputed**	
Frequency	Percent	Frequency	Percent	Frequency	Percent
64,168	96.20	466	0.70	2,072[1]	3.10

[1]This number includes 1,942 Hispanic respondents for whom donors were restricted by Hispanic-origin group (93.7% of statistically imputed cases) and 130 respondents for whom donors were unrestricted (6.3% of statistically imputed cases).

4.4.3 Hispanic-Origin (Dichotomous Indicator)

4.4.3.1 Edited Hispanic-Origin Indicator (EDQD04 and EDHOIND). Prior to creating an edited Hispanic-origin indicator, an edited version of QD04 (EDQD04) was created. If respondents indicated that they were Hispanic in response to QD03, QD04 asked them to indicate which Hispanic-origin group best describes them. If QD04's "other" category was chosen, the respondent was asked to specify an Hispanic-origin group. Unlike in previous

NHSDAs and the 1999 PAPI questionnaire, respondents had the option of selecting more than one Hispanic group in QD04, but (as in the past) the final imputed Hispanic-origin group variable was limited to one category.

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

> QD04, if it is not "other," else
>
> Hispanic-origin group recode from alpha-specify response(s),[14] if "other" was selected and a valid recode is available,[15] else
>
> missing.

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

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

If no groups were selected in QD04, EDQD04 =

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

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

---

[14] Both QD04 (Hispanic-Origin group question, see **Section 4.5.2**) and QD05/QD06 allow respondents to specify a race or Hispanic-Origin group, respectively, other than those listed in the questions, when they select the category “other”. In many cases respondents keyed in a racial category in response to the Hispanic-Origin group question (QD04) or a Hispanic-origin group in response to the race question(s) (QD05 or QD06). Thus, in checking alpha-specify responses for the race and Hispanic-origin group variables, both QD04 and QD05 were checked for each. For a detailed description of the assignment of race categories from alpha-specify responses, see **Appendix E.**

[15] In a number of cases the race and/or Hispanic-origin group specified by a respondent did not fit into the categories used by NHSDA, or the respondent did not specify a race when prompted, so not recode was available. See **Appendix E.**

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

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

missing.

4.4.3.2 Imputation-Revised Hispanic-Origin Indicator (IRHOIND). Unlike in previous years (and the 1999 PAPI sample), missing values of the CAI Hispanic-origin indicator (EDHOIND) were statistically imputed. An imputation-revised Hispanic-origin indicator for CAI called IRHOIND was created using an unweighted sequential hot-deck procedure to impute missing values of the edited Hispanic-origin indicator variable (EDHOIND). The file was serpentine sorted by design stratum, census region, race, and a random number. For a summary of nonresponse for the CAI Hispanic-origin indicator, see **Table 7**.

Table 7. Hispanic-Origin Indicator Editing and Imputation Summary

From Questionnaire		From Alpha-Specify Codes		Statistically Imputed	
Frequency	Percent	Frequency	Percent	Frequency	Percent
66,623	99.88	40	0.06	43	0.06

4.4.4 Race and Hispanicity Recodes

The imputation-revised race (IRRACE) and imputation-revised Hispanic-origin indicator (IRHOIND) variables were used to create two additional race/ethnicity variables: HISPRACE with three levels (Hispanic, non-Hispanic black, and non-Hispanic nonblack) and RACE with four levels (non-Hispanic white, non-Hispanic black, Hispanic, and non-Hispanic other). HISPRACE was used as a sorting variable in some subsequent imputations for both the PAPI and CAI variables.

Furthermore, two additional race/ethnicity variables were created to incorporate the changes to the race question in the CAI questionnaire. These variables, NEWRACE1 and NEWRACE2, provide more detail by including all of the race categories included in the question and also by indicating whether a respondent selected more than one race category. These variables indicated whether a respondent was Hispanic, based on IRHOIND, and the detailed race information was provided only for non-Hispanic respondents.

First, Hispanic respondents were assigned the NEWRACE1 and NEWRACE2 values of "Hispanic" using IRHOIND. Second, respondents with IRRACE values that were not statistically imputed were assigned values of these variables based on their responses to QD05 and QD06. Next, in those cases where NEWRACE1 and NEWRACE2 were still missing, values were assigned to respondents using the same donors as were used when the item nonrespondents' values were imputed for IRRACE. (This was done to ensure consistency between IRRACE, NEWRACE1, and NEWRACE2.) Finally, in those cases where no imputation was necessary when the levels of IRRACE were assigned, but insufficient information was provided in QD05 and/or QD06 to make assignments to the detailed levels of NEWRACE1 and NEWRACE2, a supplementary statistical imputation was performed. NEWRACE2 is a collapsed version of NEWRACE1, and an initial version of NEWRACE1 was created as follows. NEWRACE1 =

> missing, if IRHOIND = 2, QD05 = "other" and the alpha-specify response was Asian, but not one of the specific Asian categories included in NEWRACE1, else
>
> 1 (non-Hispanic white only), if IRHOIND = 2, and either white was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "white," else
>
> 2 (non-Hispanic black/African American only), if IRHOIND = 2, and either black/African American was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "black/African American," else
>
> 3 (non-Hispanic Native American or Alaska Native), if IRHOIND = 2, and either Native American was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Native American," else
>
> 4 (non-Hispanic Native Hawaiian only), if IRHOIND = 2, and either Native Hawaiian was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Native Hawaiian," else

5 (non-Hispanic Other Pacific Islander only), if IRHOIND = 2, and either Other Pacific Islander was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Other Pacific Islander," else

6 (non-Hispanic Native Hawaiian and Other Pacific Islander), if IRHOIND = 2, and both Native Hawaiian and Other Pacific Islander were selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Native Hawaiian and Other Pacific Islander," else

7 (non-Hispanic Chinese only), if IRHOIND = 2, and either Chinese was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Chinese," else

8 (non-Hispanic Filipino only), if IRHOIND = 2, and either Filipino was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Filipino," else

9 (non-Hispanic Japanese only), if IRHOIND=2, and either Japanese was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Japanese," else

10 (non-Hispanic Asian Indian only), if IRHOIND = 2, and either Asian Indian was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Asian Indian," else

11 (non-Hispanic Korean only), if IRHOIND = 2, and either Korean was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Korean," else

12 (non-Hispanic Vietnamese only), if IRHOIND = 2, and either Vietnamese was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Vietnamese," else

13 (non-Hispanic Other Asian only), if IRHOIND = 2, and either Other Asian was the only race selected in QD05, or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as "Other Asian," else

14 (non-Hispanic Asian multiple category), if IRHOIND = 2, and either more than one race was selected in QD05 where all those selected are considered "Asian," or "other" was the only race selected in QD05 and the alpha-specify response was interpreted as a combination of several Asian categories, else

15 (non-Hispanic Multiple Race), if two or more races were selected in QD05 and (a) at least one was non-Asian, (b) at least one was something other than Native Hawaiian or Other Pacific Islander, and (c) IRHOIND = 2, else

16 (Hispanic) if IRHOIND = 1.

As noted above, these assignments of NEWRACE1 values were first made for non-Hispanic respondents whose IRRACE values were not statistically imputed. Respondents whose IRRACE values were statistically imputed were assigned NEWRACE1 values from the same donors as were used for the IRRACE imputation. Those respondents who indicated "Asian" but not one of the specific Asian groups included in NEWRACE1 in an alpha-specify response for race were assigned a value of NEWRACE1 by statistically imputing a finer Asian category (12 respondents). This imputation was performed using an unweighted sequential hot-deck procedure, sorting the file by design stratum, census region, segment identification number, household type, and a random number. Donors were restricted to non-Hispanic Asian respondents with valid NEWRACE1 values. The final imputed variable was also called NEWRACE1.

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

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

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

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

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

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

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

7 (Hispanic), if NEWRACE1 = 16

4.4.5 Marital Status

4.4.5.1 Edited Marital Status (EDMARIT) . The CAI marital status question (QD07) was the same as the PAPI question used in 1999 and in past years. The base variable for creating an imputation-revised version of marital status was called EDMARIT and was created as follows. EDMARIT =

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

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

missing.

4.4.5.2 Imputation-Revised Marital Status (IRMARIT). An imputation-revised version of marital status was created using an unweighted sequential hot-deck method. For this procedure, the file was first partitioned into imputation classes based on age (to separate respondents younger than 15 in order to avoid imputing legitimate skips). The file was then serpentine sorted within classes by population density, segment identification number, race/ethnicity, age, and a random number. The imputation-revised marital status variable was called IRMARIT. See **Table 8** for a summary of item nonresponse for marital status.

Table 8. Marital Status Editing and Imputation Summary

From Questionnaire		Statistically Imputed		Legitimate Skip (≤ 14 Years Old)	
Frequency	Percent	Frequency	Percent	Frequency	Percent
53,902	80.81	36	0.05	12,768	19.14

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

4.4.6 Hispanic-Origin Group

4.4.6.1 Edited Hispanic-Origin Group (EDHOGRP). EDHOGRP, the base variable for creating an imputation-revised Hispanic-origin group variable for CAI, was created using EDQD04 and the imputation-revised Hispanic-origin indicator (IRHOIND) as follows. EDHOGRP =

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

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

missing.

4.4.6.2 Imputation-Revised Hispanic-Origin Group (IRHOGRP). Two versions were used for the imputation-revised CAI Hispanic-origin group variable: IRHOGRP and IRHOGRP3. IRHOGRP3 had seven possible values (Puerto Rican, Mexican, Cuban, Central or South American,[16] Caribbean Islander, or Other-Hispanic) and was created using an unweighted sequential hot-deck procedure to impute missing values of EDHOGRP. Imputation classes were created based on Hispanic origin (to avoid imputing legitimate skips from non-Hispanic respondents) and race, and the file was serpentine sorted within classes by State, design stratum, segment identification number, race, and a random number. After IRHOGRP3 was imputed, IRHOGRP was created by recoding IRHOGRP3 into four groups: Puerto Rican, Mexican, Cuban and "other," where "other" includes Other-Hispanic, Central or South American, and Caribbean Islander. For a summary of item nonresponse for Hispanic-origin group, see **Table 9**.

Table 9. Hispanic-Origin Group Editing and Imputation Summary

From Questionnaire		From Alpha-Specify Codes		Statistically Imputed		Legitimate Skip (Not Hispanic)	
Frequency	Percent	Frequency	Percent	Frequency	Percent	Frequency	Percent
7,707	11.55	637	0.95	137[17]	0.21	58,225	87.29

[16] In PAPI (both for 1999 and in past years) there were seven levels of IRHOGRP2; in the 1999 CAI, however, Central American and South American were collapsed into one category in IRHOGRP3 due to changes in the questionnaire.

[17] This includes 42 respondents with donors restricted by race (30.7% of statistically imputed cases) and 95 whose donors were unrestricted (69.3% of statistically imputed cases).

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). An imputation-revised version of EDEDUC was created using an unweighted sequential hot-deck procedure to impute missing values. The file was serpentine sorted by design stratum, population density, segment identification number, marital status, age, race/ethnicity, and a random number. The imputation-revised version of this variable is called IREDUC. For a summary of nonresponse for highest grade completed, see **Table 10**.

Table 10. Highest Grade Completed Nonresponse Summary

From Questionnaire		Statistically Imputed	
Frequency	Percent	Frequency	Percent
66,680	99.96	26	0.04

4.4.7.3 Education Recode. EDUCCAT2, a recoded education variable, was created using the imputation-revised highest grade completed variable (IREDUC) for both PAPI and CAI cases. 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).

This page intentionally left blank.

5. Noncore Demographics

5.1 Introduction

In past years and in the PAPI supplemental sample of the 1999 NHSDA, no noncore demographic variables except income and health insurance were imputed. Current work status was a core demographic variable determined from a single question. In CAI, however, current work status was a noncore variable determined from multiple questions and differed substantially from the single PAPI question. Instead of a single question asking the respondent to describe their "current" work status, several questions were asked regarding the respondent's work situation in the week prior to the interview, and also more generally if that week was not typical. Furthermore, in PAPI and past NHSDAs the work status question was asked of all respondents, but the CAI work status questions were asked only of respondents aged 15 or older.

5.2 Current Employment Status

5.2.1 Edited Employment Status

5.2.1.1 JOBSTAT. The variable used to summarize the respondent's current work situation (in the week prior to the interview) was the edited variable JOBSTAT, which combined the information from questions QD26 to QD31b, the primary work status questions in the CAI questionnaire. The categories for JOBSTAT are shown in **Exhibit 2**. Many of the JOBSTAT categories were created using the alpha-specify responses[18] to two questions (QD30 and QD31a) regarding why the respondent did not work at a job or business on the week before the interview. These responses were coded in a dictionary, the details of which are discussed in **Appendix F**. (The two questions and the listed responses are also given in Appendix F.) Details about the creation of JOBSTAT and the mapping of questionnaire responses (including the dictionary codes in Appendix F) are discussed in a separate document (Kroutil, 2001a).

In past years (and in the 1999 PAPI), missing values for a 12-level current work status variables were imputed. The imputation-revised variable was then recoded to obtain an employment status variable called EMPSTAT2 with five levels: 12 to 17 year old, full-time

[18]"Alpha-specify responses" are written responses entered by the interviewer when the listed responses in the given question were judged to be inapplicable (see **Section 3.2**).

Exhibit 2. Categories of JOBSTAT

#	Employment Situation	#	Employment Situation
1	Worked at full-time job, past week	191	Has part-time job, reason for not working unclear
2	Worked at part-time job, past week	192	Has job, did not want to work past week
3	Has job but out: vacation/sick/temp absence	193	Has job during school year, no further information
4	Has job but out: layoff, looking for work	199	Has job, no further information
5	Has job but out: layoff, not looking for work	201	Volunteer worker
6	Has job but out: waiting to report to new job	202	Does not need to work
7	Has job but out: self-employed, no business past week	203	Does not want to work
8	Has job but out: in school/training	204	Cannot work, reason unspecified
9	No job: unemployed/layoff, looking for work	205	Not eligible/not allowed to work
10	No job: layoff, not looking for work	206	No job: family responsibilities
11	No job: keeping house full time	207	No job: starting/finished school
12	No job: in school/training	208	Student/youth, looking for work
13	No job: retired	209	No job: substance abuse issues
14	No job: disabled for work	211	No job: income restrictions
101	Seasonal worker	212	No job: literacy, language, learning disabilities, etc.
102	Not scheduled/temp/on-call worker	213	Not working due to legal issues
103	Babysitter	290	Unemployed, no further information
104	Full-time during school year	291	Doesn't/never worked, reason unspecified
105	Part-time during school year	299	Other, not in labor force
106	Missionary/religious worker	900	Work status unclear
190	Has full-time job, reason for not working unclear	Remaining codes in the 900 series have their standard meanings in the NHSDA: Bad data (985), Legitimate skip (991), Don't know (994), Refused (997), Blank (998), etc.	

work, part-time work, unemployed, and other. JOBSTAT had many more levels than just 12, however, and missing values in JOBSTAT were not replaced with imputed values. No intermediate 12-level variable was created in the 1999 CAI sample, so a 5-level employment variable similar to EMPSTAT2 was created (called EMPSTAT3), and the missing values in EMPSTAT3 were imputed directly.

5.2.1.2 EDEMP. The base variable used to create an imputation-revised employment status variable for CAI, EDEMP, was derived using JOBSTAT in the following manner. EDEMP =

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

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

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

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

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

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

missing.

5.2.2 Imputation-Revised Employment Status

Missing values in the edited employment status variable EDEMP were replaced with imputed values using an unweighted sequential hot-deck procedure. Prior to imputation, respondents aged 12 to 17 were separated from the rest of the file to avoid imputing EMPSTAT3 values of five; these cases were added back into the file following imputation. Note that all respondents aged 12 to 17 were assigned to a single category of the final employment status variable. The file used for imputation was serpentine sorted by design stratum, highest grade completed, race/ethnicity, gender, age, and a random number. Missing values were imputed as follows:

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

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

For a summary of item nonresponse for CAI employment status, see **Table 11**.

Table 11. Employment Status Editing and Imputation Summary

From Questionnaire		**Statistically Imputed**		**12 to 17 Years Old**	
Frequency	Percent	Frequency	Percent	Frequency	Percent
41,034	61.52	315	0.47	25,357	38.01

6. CAI Drug Imputations

The major changes introduced in the imputation procedures for the drug use variables in the CAI sample of the 1999 NHSDA have already been alluded to in **Section 3.2.** Missing values for more variables were imputed in the 1999 CAI sample than in the PAPI for1999 or in previous NHSDAs, and more sophisticated techniques were used. As stated in Section 3.2, the unweighted sequential hot deck was mostly abandoned[19] in favor of a new procedure called predictive mean neighborhoods (PMN), a combination of weighted regression and nearest neighbor hot-deck imputation, where the hot deck is random whenever possible.[20]

This chapter describes how the PMN technique was applied to the drug use variables. In some cases, imputations were required because the respondent did not answer a given question. However, other responses were altered in the editing process due to inconsistencies. In these cases, the original response was either set to missing, or in the case of recency of use, a specific recency was edited to a more general recency that was consistent with other responses, and determination of the specific recency was left to imputation. For example, a recency-of-use response might be edited to past year usage, where past month versus past-year-but-not-past-month use is determined by imputation. The aforementioned editing processes are summarized by Kroutil (2001a).

The models for these imputations, which are described in detail in the following sections, were either binomial or multinomial weighted logistic models, or weighted multiple linear regression models with the response variable appropriately transformed. Using the PMN technique, the predicted means from these models were used to determine neighborhoods, from which donors were randomly selected for the final assignment of imputed values. (If no donors were available within a very small distance of the recipient's predicted mean, then the donor with the closest predicted mean was chosen.) The neighborhoods were created based on a single predicted mean (a univariate predictive mean neighborhood [UPMN]), or using several predicted means at once (a multivariate predictive mean neighborhood [MPMN]). Even if the neighborhood is constructed from a univariate predicted mean, the assignment of imputed values may be either univariate or multivariate. The members of the neighborhood were restricted to satisfy two types of constraints: "logical constraints" and "likeness constraints." Constraints that

---

[19]A form of the unweighted sequential hot deck was used to determine provisional lifetime usage indicators.

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

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

Because drug use is highly correlated with age, and to facilitate easier implementation of the imputation procedures, the model building and final assignment of imputed values for all drug use variables were performed separately within three distinct age groups: 12 to 17 year olds, 18 to 25 year olds, and persons 26 years of age or older.[21]

Although statistical imputation of the drug use variables could not proceed separately within each State due to insufficient pools of donors, information about the State of residence of each respondent is incorporated in the modeling and hot-deck steps in the CAI sample. States were classified into three drug usage categories: States with high usage of a given drug were placed in one category, States with medium usage into another, and the remainder into a third category. Respondents were then assigned values for a three-level"State rank" variable, depending on their State of residence. The indicator variables resulting from this categorical State rank variable were used as covariates in the imputation models. In addition, for all of the drug use measures except lifetime usage,[22] 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.

The CAI sample of the 1999 NHSDA had different drugs and different measures of drug use than in the PAPI supplemental sample of the 1999 NHSDA (or in past years' PAPIs). **Exhibit 3** summarizes the drugs and drug use measures that were imputed and whether the imputations were univariate or multivariate. If no character is present in the box, then no information regarding that particular drug use measure was available for the given drug.

---

[21]Modeling was done separately within each of the three age groups regardless of the response variable. However, to ease computational burden, the 12- to 17-year-old age group was split into males and females in the assignment step for lifetime usage of each drug.

[22]The lifetime usage imputation was performed at an early stage in the imputation process. The UPMN programs were not sufficiently developed at the time of the lifetime usage imputation to allow for imputation within state rank groups.

Exhibit 3. 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 recency of use, 12-month frequency of use where applicable, 30-day frequency of use where applicable, and the 30-day binge drink frequency variable (alcohol only); multivariate assignment of imputed values across measures.

✕✕ Multivariate neighborhood across recency of use, 12-month frequency of use where applicable, and 30-day frequency of use where applicable; multivariate assignment of imputed values across these measures, and across certain drugs (e.g., see **Sections 6.3.1.7.1**, **6.3.1.7.2**, **6.3.1.7.3**, and **6.3.2.7**).

✓✕ Univariate neighborhood and multivariate assignment of imputed values (see **Sections 6.2.7.1**, **6.2.7.3**, **6.4.1.7.1**, **6.4.1.7.2**, and **6.4.1.7.3**).

†✕ Univariate neighborhood (incorporates predicted values for both cocaine and crack (see **Section 6.2.7.2**) and multivariate assignment of imputed values.

[1]Includes chewing tobacco and snuff.

[2]Includes LSD and PCP.

[3]Includes methamphetamines.

6.1 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 be used as covariates for models fitted later in the sequence. Because the gate questions are the basis for all subsequent drug data, the imputation of missing values for lifetime drug use for all drugs must precede imputations of all other drug use measures. These lifetime use indicators are temporary in the sense that they are manifested within the drug recency and frequency-of-use variables, but are not delivered themselves. The hierarchy of models for drugs for the lifetime usage models is discussed in **Section 6.2**.

Once all the lifetime usage indicators had been determined, the imputations of the remaining measures could proceed. As indicated in **Exhibit 3**, a multivariate imputation was implemented across the measures within each drug for recency of use, 12-month frequency of use, 30-day frequency of use, and binge drink frequency (alcohol only). For a given drug, recency of use was included in the model for frequency of use, 12-month frequency of use was included in the model for 30-day frequency, and 30-day frequency of use of alcohol was included in the model for the binge drink frequency variable. Finally, age at first use must be consistent (in a number of ways) with the other measures (see **Section 6.4**). Hence, age at first use was imputed after the imputation for the other measures was completed.[23] The following sections describe the imputation procedures for each drug use measure.

6.2 Imputing Lifetime Drug Use Indicators

Unlike previous NHSDAs, the 1999 NHSDA CAI implemented automatic routing through the questionnaire. Using a series of gate questions, the instrument asked the respondent whether he or she had ever used a number of drugs in his or her lifetime. Based on the response to each gate question, the instrument either routed the respondent through the current drug module or skipped him or her to the next module. Thus, the respondent was not necessarily required to answer all questions in the questionnaire. The respondent could skip a module if he or she either indicated nonusage of the drug in the gate question or did not answer the gate question. Therefore, the gate question response was key to the range of responses available for subsequent questions in each module.

---

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

6.2.1 Hierarchy of Drugs

The first step in the imputation of lifetime indicators was to determine the order in which the drugs would be modeled (i.e., the "drug hierarchy" discussed in detail in **Appendix D**). For a particular drug, it was expected that indications of lifetime use of other drugs would be strong predictors of lifetime use of that drug. Hence, drugs expected to be highly correlated with the lifetime use of other drugs were placed later in the sequence. Note that the lifetime usage indicators, when used as predictors, were only provisional because the final imputation of lifetime usage indicators was not implemented until the lifetime usage modeling was completed for all drugs. The order in which the lifetime indicators of use were imputed is shown in **Exhibit 4**.

6.2.2 Setup for Model Building and Hot-Deck Assignment

Once the hierarchy of drugs was established, the next step was to define respondents, nonrespondents, and the item response mechanism. As stated earlier, imputations for all drug use measures were conducted separately within the three age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents 26 years of age or older. For an individual to be considered a lifetime use item respondent, he or she must have complete data within each age group for all of the drug module gate questions: cigarettes, cigars, chewing tobacco, snuff, pipes, alcohol, marijuana, cocaine, crack, heroin, inhalants, LSD, PCP, hallucinogens other than LSD and PCP, analgesics, tranquilizers, methamphetamines, stimulants other than methamphetamines, and sedatives. Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. The predicted probability P (survey respondent is an item respondent | respondent is a lifetime user) was determined for each item respondent from this model, the inverse of which was multiplied by the respondent's weight. Note that because item respondents were defined across all drugs, this adjustment was only computed once per age group and then used in the modeling of lifetime use for all drugs. The item response propensity model is a special case of the generalized exponential model (GEM), which is described in greater detail in **Appendix C**.

For certain categories of drugs, multiple gate questions within a drug module were used to assess lifetime use or nonuse of the overall group of drugs within that module (e.g., LSD, PCP, and a number of other substances within the drug module for hallucinogens were used to assess usage of hallucinogens). For these drug groups, if any of the gate questions were answered "yes" (i.e., the respondent indicated using the drug once or more in his or her lifetime), then the lifetime use indicator for the overall drug group was set to "yes." For example, to assess lifetime

Exhibit 4. Lifetime Indication of Use ("Gate") Questions for CAI (in Order of Imputation)

Drug	**Question(s)**
Cigarettes	CG01
Smokeless Tobacco[1]	CG17, CG25a, CG25b
Cigars	CG34
Pipes	CG42
Alcohol	AL01
Inhalants	IN01a, IN01b, IN01c, IN01d, IN01e, IN01f, IN01g, IN01h, IN01i, IN01j, IN01l
Marijuana	MJ01
Hallucinogens[2]	LS01a, LS01b, LS01c, LS01d, LS01e, LS01f, LS01h
Pain Relievers	PR01, PR02, PR03, PR04, PR05
Tranquilizers	TR01, TR02, TR03, TR04, TR05
Stimulants[3]	ST01, ST02, ST03, ST04, ST05
Sedatives	SV01, SV02, SV03, SV04, SV05
Cocaine	CC01
Crack	CK01
Heroin	HE01

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

use of the overall drug group "inhalants," the respondent was asked if he or she had ever, even once, inhaled any of the following with the intention of getting high: (1) amyl nitrite, "poppers," locker room odorizers, or "rush"; (2) correction fluid, degreaser, or cleaning fluid; (3) gasoline or lighter fluid; (4) glue, shoe polish, or toluene; (5) halothane, ether, or other anesthetics; (6) lacquer thinner or other paint solvents; (7) lighter gases, such as butane or propane; (8) nitrous oxide or whippets; (9) spray paints; and (10) any other aerosol spray. If the response to any of 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, analgesics, tranquilizers, stimulants, sedatives, and smokeless tobacco. To be considered a nonrespondent of a drug module with multiple gate questions, the respondent had to answer "no" to all of the gate questions. If none of the gate questions in a drug module was answered affirmatively, but some of the gate questions were unanswered, the individual was considered a nonrespondent for that module.

6.2.3 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. The parameters for the models were determined using backward elimination in logistic regression. The predictors considered in each model included continuous age, age squared, age cubed, race/ethnicity, gender, lifetime use of drugs already imputed, 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, age squared, age cubed, race/ethnicity, and gender. For age groups 18 years of age or older, the variables marital status, education, and employment status were also included. For a complete summary of the lifetime use imputation models, see **Appendix G**.

6.2.4 Computation of Predicted Means, Assignment of Provisional Imputed Values, and Creation of Univariate Predictive Mean Neighborhoods

Using the parameters from the probability of lifetime use 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 an unweighted sequential hot deck. In other words, the file was sorted by the predicted probabilities of lifetime use, and missing values of the lifetime use indicator were replaced with the value from the previous item respondent. (As with the response propensity adjustment and modeling, the sorting and assignment of imputed values were performed separately within each of the three age groups.) To avoid bias associated with the direction of the sort and to avoid missing values at either end of the file, the file was sorted in both ascending and descending orders by predicted probability of lifetime use. A value was then borrowed by randomly selecting the preceding donor. (In a descending sort, the preceding donor is the same as the succeeding donor in an ascending sort.) If the nonrespondent was the first or last record on the file, the value was forced to be borrowed in the direction in which a donor existed. The resulting provisional imputation-revised value was then used as a covariate in the model for the subsequent drug in the sequence.

Once the modeling of lifetime use for all drugs was completed, the final assignment of imputed values was implemented using the UPMN technique described in **Appendix D**.[24] This procedure defines a "neighborhood" of respondents by requiring that these respondents' predicted mean values be within a certain distance, denoted by delta, of the nonrespondent's value. In this case, the predicted mean was the predicted probability of lifetime use from the model. The value of delta varied depending on the value of the predicted mean, which in this case was the predicted probability of lifetime use. This allowed a looser delta for predicted probabilities close to 0.5, and a tighter delta for predicted probabilities close to zero or 1. The range of values for delta across various predicted probabilities is given in **Exhibit 5**.

Exhibit 5. Values of Delta for Various Predicted Probabilities of Lifetime Use

Predicted Probability (p)	Delta
$0.4 < p < 0.6$	0.05
$0.25 < p \leq 0.4$; $0.6 \leq p < 0.75$	0.04
$0.15 < p \leq 0.25$; $0.75 \leq p < 0.85$	0.03
$0.05 < p \leq 0.15$; $0.85 \leq p < 0.95$	0.02
$p \leq 0.05$; $p \geq 0.95$	0.01

6.2.5 Assignment of Final Imputed Values

For the final assignment of an imputed value for a given drug, a value from one member of the neighborhood was randomly selected to replace the missing response. Separate assignments were performed within each of the three age groups, subject to the constraints described in the next section. For details on the UPMN methodology, see **Appendix D**.

6.2.6 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

[24]The UPMN methodology could have been used in lieu of the provisional assignment of a value using the unweighted sequential hot-deck methodology. However, this procedure was not fully developed at the time of the lifetime indicator of use modeling.

recipients as possible. As with all other drug use measures, neighborhoods for lifetime use indicators were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, and 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. A small delta could also be considered a likeness constraint, which could be loosened by enlarging delta. This was never done, however, with the lifetime usage indicators.

No logical constraints were placed on the neighborhoods for any of the lifetime usage indicators. Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. Even in those cases, however, the imputation was carried out so that no logical constraints were necessary, as discussed in **Section 6.2.7**.

6.2.7 Multivariate Assignments

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

6.2.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff). Many respondents who indicated lifetime use of smokeless tobacco seemed to be confused regarding the difference between chewing tobacco ("chew") and snuff, as was demonstrated by their responses to questions regarding specific brands. For example, many respondents who indicated use of chewing tobacco entered a snuff brand, such as "Copenhagen," when asked about the specific brand of chew they used. As a result, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted, rather than individual models for chew and snuff. The nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted probability of lifetime use. Missing values for chew and/or snuff were replaced with the values from a donor within this neighborhood. For individuals missing the lifetime usage indicator for either chew or snuff but not both, only the missing value was replaced. However, for individuals missing both chew and snuff, both lifetime usage indicators were replaced by values from the same donor. No logical constraints were necessary in the assignment step because chew and snuff were assigned values independently, then combined at the end to form a final lifetime usage indicator for smokeless tobacco.

6.2.7.2 Cocaine and Crack. Because cocaine and crack were in distinct modules in the CAI questionnaire, separate models were fit for the two substances. However, crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. This neighborhood was defined in terms of the deltas given in **Exhibit 5**, 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. Note that it would not be possible for a respondent to be missing a value for cocaine, but not crack, because a crack user is by definition also a cocaine user. For this reason, no logical constraints were necessary.

6.2.7.3 Hallucinogens (LSD, PCP, and Other Hallucinogens) and Stimulants (Methamphetamines and Other Stimulants). The modules for both hallucinogens and stimulants included multiple gate questions (called subgate questions), and some of the substances referred to in the subgate questions were of interest in their own right. For hallucinogens, there was interest in the usage of LSD and PCP; for stimulants, there was interest in the usage of methamphetamines. Predicted probabilities were calculated for the larger groups of substances known as hallucinogens and stimulants, and these probabilities were used to determine neighborhoods for each group of drugs. An "other" category was created by combining all the other subgate questions except the ones of special interest. In the final assignment step, lifetime usage indicators were assigned for LSD, PCP, and "other" for hallucinogens, and for methamphetamines and "other" for stimulants. The final lifetime usage indicators for hallucinogens and stimulants were created by combining the constituent parts, including the "other" group of substances.

6.2.7.3.1 Hallucinogens. The lifetime usage indicator for "other hallucinogens" was created using the lifetime usage information from all the hallucinogens' subgate questions except LSD and PCP. Note that if a respondent was a user of at least one of the other hallucinogens, then he or she was considered a user of other hallucinogens, even if some of the other hallucinogens' subgates were unanswered. A missing value for other hallucinogens arose if at least one of the other hallucinogens' subgate questions was unanswered, and all the other hallucinogens' subgate questions that were answered had a negative response. Using the neighborhood created from the hallucinogens' predicted probability of lifetime use, missing values for LSD and/or PCP and/or other hallucinogens were replaced with the values from a donor within this neighborhood. For individuals missing a lifetime usage indicator for either LSD and/or PCP and/or other hallucinogens, only the missing value(s) was (were) replaced. For individuals missing two or more of these lifetime usage indicators, the missing values were replaced by values from the same donor. As with smokeless tobacco, the subcategories for hallucinogens were assigned values separately, making logical constraints unnecessary. As a final step, a lifetime usage indicator for all hallucinogens was created by combining the lifetime usage indicators for the three subgroups.

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

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

In the CAI sample of the 1999 NHSDA, the drug use measures' recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and (for alcohol) 30-day binge drinking frequency[25] were modeled separately for each drug. Provisional values replaced missing values for use in subsequent models, where necessary, using the UPMN methodology described in **Appendix D**. After having modeled all of the drug measures, the MPMN methodology (also described in Appendix D) was employed to determine final imputed values using the predicted

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

values from these models. If no donor could be found using the MPMN technique, even after loosening likeness constraints, UPMN values were used as final imputed values.

An error was detected in the multivariate imputation of recency and frequency of use, requiring the imputation of missing values in recency and frequency of use to be redone. For most drugs, the updated numbers are nearly identical to the old ones, except for estimates of alcohol, marijuana, and inhalant use. For these three drugs, as well as the "any illicit drug use" measure that is subsequently calculated, the differences are noticeable, especially for inhalants. Details about the error, and how the error was fixed, are described in **Appendix H**.

6.3.1 Recency of Use

6.3.1.1 Hierarchy of Drugs. A complete drug hierarchy, as described in **Appendix D**, was not required for recency of use because only cigarettes, alcohol, and marijuana recencies were used as covariates in models for subsequent drugs. This was due to difficulties that would arise if too many covariates were included in the polytomous logistic models. (Lifetime usage indicators of other drugs were included instead of recency-of-use indicators.) The cigarettes' recency was modeled first, and the predicted probability of past month use was used to determine provisional values[26] used in the models for cigars, smokeless tobacco, pipes, and alcohol recency of use. Alcohol and marijuana followed the tobacco products in the sequence. The order in which the remainder of drugs were modeled depended upon the availability of logically edited variables and the amount of time required to fit each model.

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

---

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

Before modeling, the respondents' weights were adjusted so that they represented all lifetime users. Because item respondents were defined at the drug level, these adjustments were made separately for each drug (and within the three age groups). Adjustments were made using an item response propensity model (see **Appendix C** for the more general GEM), and covariates included a categorical age variable, race, gender, census region, an MSA[27] indicator, and provisional cigarette, alcohol, and marijuana recencies (for those drugs following cigarettes, alcohol, and marijuana).

6.3.1.3 Sequential Model Building. Using the adjusted weights, the probability of selecting each recency-of-use category was modeled within each age group using polytomous logistic regression. The predictors included in the models were age; age squared; gender; race; first-order interactions of age, age squared, gender, and race; marital status; education; employment status[28]; census region; MSA indicator; provisional 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.[29] Because interest was only in the estimation of the predicted mean, and not in the parameter estimates (by themselves) or their standard errors, no model selection was attempted. For a summary of the variables included in each drug model, see **Appendix G**.

6.3.1.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods. As mentioned previously, provisional recency-of-use values for cigarettes, alcohol, and marijuana were created following the modeling of each of these drugs so that they could be used as covariates in subsequent models. This was necessary because, although item respondents had complete data on a particular drug, it was often the case that an item respondent for one drug was an item nonrespondent for another. Provisional recency-of-use values were also created for the remainder of the drugs so that they could be included as covariates in the modeling of other measures within the multivariate set (i.e., 12-month frequency and/or 30-day frequency of use). 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

---

[27]MSA refers to a metropolitan statistical area, as defined by the Office of Management and Budget (OMB).

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

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

nonrespondents, using the parameters from the polytomous logistic model. The predicted probabilities from the recency models were used to assign provisional values using the UPMN imputation method described in **Appendix D**. A vector of predicted probabilities for each respondent was created by the polytomous logistic regression model. Because only a single predicted mean was used to determine the neighborhood when determining provisional values, not all the predicted probabilities from the model were used.[30] Because past month use was the most critical measure of recency of drug use, the neighborhoods were defined based on the probability of past month use. If possible, provisional donors were chosen with predicted means within the delta[31] of the recipient, where the value of delta varied depending on the value of the predicted means, which in this case were predicted probabilities of past month use.[32] In particular, delta was defined as 5% of the predicted probability if the probability was less than 0.5, and 5% of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5, and a tighter delta for predicted probabilities close to zero or 1. If no donors were available with predicted means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predicted mean was chosen.

6.3.1.5 Assignment of Provisional Imputed Values. Separate assignments of provisional values were performed within each of the three age groups, subject to the constraints described in the next section. The final recency-of-use imputations were multivariate across drug measures and are further described in **Section 6.3.5**.

6.3.1.6 Constraints on Univariate Predictive Mean Neighborhoods. As stated in the lifetime usage section, a UPMN neighborhood can be restricted by logical constraints (which cannot be loosened) and by likeness constraints (which can be loosened) to make candidate donors in the neighborhood as similar to recipients as possible. As with all other drug use measures, neighborhoods for recency of use were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, or 26 or older). Models were

---

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

[31]"Delta" refers to the value that defines the neighborhood of donors that are "close" to the item nonrespondent. The difference between the predicted mean of the item nonrespondent and the predicted means of the item respondents in the neighborhood must be less than delta. See **Appendix D** for more details.

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

built separately within these three groups, so this likeness constraint was never loosened. A small delta could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. As previously stated, if no donors could be found in the delta as defined in **Section 6.3.1.4**, the neighborhood was abandoned, and the donor with the predicted mean closest to the recipient was chosen.[33] If possible, donors and recipients were required to be from States with the same level of usage of a given drug (the State rank, as defined in the introduction to this chapter), where the level of usage was defined in terms of the proportion of a given State's residents who had used a given drug in the past month. If insufficient donors were available within these constraints, they were loosened in the following order: (1) the neighborhood was abandoned, and the donor with the closest predicted mean was chosen; (2) donors and recipients were no longer required to be from States with similar usage levels. **Appendix I** gives a summary of how many respondents had values imputed under various constraints.

Logical constraints were placed on the neighborhoods in those cases where a general recency category was available for a respondent and imputation was required to determine the specific recency categories. The general recency categories that appeared, and the restrictions on possible donors that did not involve an interview date, are given in **Exhibit 6**. As indicated in the exhibit, an additional logical constraint was applied only to tobacco products: If the respondent's age at first use was within 2 years of his or her current age, it would be impossible for a respondent to have last used the substance more than 3 years ago. Hence, under these circumstances, the donors were limited to have used within the past 3 years. Such a logical constraint would not be useful for nontobacco products because the recency categories for lifetime use but not past 3 year use and for past 3 year use but not past year use were combined into a single category for lifetime use but not past year use. Additional logical constraints, not listed in Exhibit 6, limited the possible recencies that could be assigned based on the respondent's current age, the time between the interview date and the birth date, the time between the interview date and the month of first use, and any nonmissing frequency of use information. A complete list of missingness patterns across recency and frequency of use (including patterns with general recency categories), and the logical constraints that correspond to those missingness patterns, is given in **Appendix J**. See **Section 6.3.5** for a discussion of the multivariate imputation of recency and frequency of use.

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

Exhibit 6. Logical Constraints on Univariate Predictive Mean Neighborhoods (Not Involving Interview Date) When a General Recency Category Was Given

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

Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. Those cases are discussed in detail in the next section (**Section 6.3.1.7**).

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

6.3.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff). For reasons discussed in **Section 6.2.7.1**, one model for smokeless tobacco (a combination of the chew and snuff responses) was fit rather than individual models for chew and snuff. The nearest neighbor hot-deck neighborhood was then based on the predicted probability of past month use of smokeless tobacco. Missing recency-of-use values for chew and/or snuff were replaced with the (provisional) values from a donor within this neighborhood. At this stage in the process, lifetime use or nonuse of either chew or snuff was considered known (employing information from the lifetime usage imputation). For lifetime users of chew or snuff who were missing some or all of their recency-of-use information[34] for either chew or snuff but not both, only the missing specific recency-of-use values were replaced. However, for individuals missing recency-of-use information for both chew and snuff (given that the respondent was known or was imputed to be a chew user and a snuff user), values for both were obtained from the same donor. The provisional recency of use for smokeless tobacco was obtained by combining the recency-of-use information from snuff and chew.

Unlike recency of use, separate models for snuff and chew were built for 30-day frequency of use. The predicted means from these models were conditioned on past month use. In the 30-day frequency of use imputations, which are discussed in **Section 6.3.3.3**, the predicted means used to form the neighborhoods were conditioned on lifetime usage rather than past month usage. Because the 30-day frequency models gave predicted means conditioned on past month use, it was necessary to determine the probability of past month use given lifetime use, which can be obtained from the recency models. Because the 30-day frequencies for snuff and chew could not be combined, recency-of-use models were built for snuff and chewing tobacco separately, where the response was past month use versus not past month use. (This was in addition to the regular recency-of-use model that was built for smokeless tobacco.) See **Section 6.3.3.3** for more details. The covariates used in the models are the given in **Appendix G**.

6.3.1.7.2 Cocaine and Crack. Even though cocaine and crack are in distinct modules in the CAI questionnaire, a recency model was only fit for cocaine. Crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood. As with the other drugs, the neighborhood was defined in terms of delta, where the value of delta varied depending on the value of the predicted means, which in this case were predicted probabilities of past month use of cocaine. In particular, delta was defined as 5% of the predicted probability if the probability was less than 0.5, and 5% of 1 minus the predicted probability if the probability was greater than 0.5. As with smokeless tobacco, use or nonuse of crack was considered known (using information from the lifetime imputations). Hence, as a logical constraint, users of crack with incomplete recency information required donors who were also crack users. Moreover, if the cocaine recency was not missing, the donated crack recency could not be more recent than the

[34]For respondents missing all of their recency information, the only known information is that they were lifetime users (either from their survey response or from imputation). For respondents missing some of their recency information, they might have been assigned a general recency category (outlined in **Exhibit 6**), and specific recency values needed to be imputed.

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

6.3.1.7.3 Hallucinogens (LSD, PCP, and Other Hallucinogens) and Stimulants (Methamphetamines and Other Stimulants). As stated in **Section 6.2.7.3**, the modules for hallucinogens and stimulants included subgate questions referring to substances that were of interest in their own right. For hallucinogens, there was interest in the usage of LSD and PCP; for stimulants, there was interest in the usage of methamphetamines. Recency-of-use information for both hallucinogens and stimulants was used in subsequent models; LSD, PCP, and methamphetamines recencies of use were not used. Hence, obtaining provisional values for the recency of use of the substances corresponding to the subgate questions was less crucial. The imputation of missing values for these substances was still carried forward, however, in case the MPMN technique could not be used to obtain final imputed values, and the UPMN values were required as a fallback.

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

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

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

6.3.2 12-Month Frequency of Use

6.3.2.1 Hierarchy of Drugs. The imputation of the 12-month frequency-of-use variables was not sequential.[35] However, the imputation of recency of use had to be completed for the drug of interest prior to imputation of the 12-month frequency of use because this measure was included in the 12-month frequency of use model.[36] (Recency-of-use variables for other drugs were also included in the model for a particular drug [see **Section 6.3.2.3**].) 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 3**).

6.3.2.2 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

---

[35]Because indicators of usage of other drugs were included in the 12-month frequency-of-use model for a given drug (and the 12-month frequency was hierarchically related to lifetime usage and recency of use), including the 12-month frequency of use for other drugs as covariates in the model would introduce complications, with little gain.

[36]Because the 12-month frequency models were limited to past year users, only two recency categories could result: past month use and past year but not past month use. Hence, recency of use for the drug being modeled was represented by a single indicator variable representing these two categories.

respondent had to have a valid response to all of the applicable measures for the drug of interest. The response propensity adjustment modeling included age, race, gender, census region, an MSA indicator, and (where available) recencies of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives as predictors.[37]

6.3.2.3 Model Building. As was apparent from the previous section, only past year users of the drug of interest were used to build the 12-month frequency-of-use model. The (untransformed) response variable of interest in the 12-month frequency-of-use models for most respondents was the proportion of the days in a full year (365.25) on which a respondent used a particular drug. For example, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/365.25. Some respondents, however, started using the drug within the past year. If they responded to the month at first use question, the difference between the month at first use and the date of the interview indicated the total time period during which they could have used.[38] If the date of the interview was July 10th, for example, and the month of first use was March, the maximum period during which the respondent could have used is the number of days between March 1st and July 10th, or 101. Thus, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/101 instead of 100/365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

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

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

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

---

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

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

where P_i is defined as the proportion of days in the past year in which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[39] Using the adjusted weights, a linear univariate regression model was then fit for the log-transformed variable Y_i, within each age group.

The recency-of-use modeling and 12-month frequency-of-use modeling were performed at almost the same time. As stated earlier, it was required that the recency of use for the drug in question be available so that the past month use indicator could be used as a covariate in the 12-month frequency-of-use model. However, only some of the recency-of-use variables for other drugs were available when the 12-month frequency-of-use was being modeled for a particular drug. For drugs for which the recency of use was not yet modeled, the lifetime indication of use served as a surrogate for the recency-of-use indicators. To control for State variations in drug use, the State rank groups defined for the recency-of-use imputations were included as covariates in the 12-month frequency-of-use models.[40] Thus, the models included age; age squared; age cubed; gender; race; State rank (based on past month prevalence of the drug); marital status; employment; educational level[41]; census region; an MSA indicator; (where available) the provisional recencies of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; as well as first-order interactions of age, age squared, age cubed, gender, and race.[42] Because interest focused only on the estimation of the predicted mean, and not on the parameter estimates (by themselves) or their standard errors, no model selection was attempted. Predicted 12-month frequencies of use were defined by back-transforming the resulting predicted values. For a complete summary of the 12-month frequency-of-use models, see **Appendix G**.

---

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

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

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

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

The predicted mean that comes out of the 12-month frequency-of-use model is a logit of the proportion of the year used. This logit was transformed back into a proportion for use as the variable from which the neighborhoods were created. If the proportion is interpreted as a probability, with the probability of use on a given day in the year conditional on past year use,[43] this probability could be multiplied by the probability of past year use to make the predicted mean conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it is not known whether they are past year users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

6.3.2.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods. Within a given drug, predicted means from the 12-month frequency-of-use models were computed for both item respondents and item nonrespondents using the parameters from the regression model. The logits were converted back to proportions, which were in turn multiplied by the probability of past year use to make the predicted mean conditional on lifetime use. Using the UPMN methodology described in **Appendix D**, neighborhoods were defined based on these predicted means. If possible, provisional donors were chosen with predicted means within delta[44] of the recipient, where the value of delta varied depending on the value of the predicted means, which in this case were predicted proportions of the year used. In particular, delta was defined as 5% of the predicted proportion if the proportion was less than 0.5, and 5% of 1 minus the predicted proportion if it was greater than 0.5. This allowed a looser delta for predicted proportions close to 0.5, and a tighter delta for predicted proportions close to zero or 1. As with recency of use, if no donors were available with predicted means within delta of the recipient, the neighborhood was abandoned and the donor with the closest predicted mean was chosen.[45]

---

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

[44]"Delta" refers to the value that defines the neighborhood of donors that are "close" to the item nonrespondent. The difference between the predicted mean of the item nonrespondent and the predicted means of the item respondents in the neighborhood must be less than delta. See **Appendix D** for more details.

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

6.3.2.5 Assignment of Provisional Imputed Values. For all drug use measures except 12-month frequency, the observed value of interest was donated directly to the recipient. However, because donors and recipients could potentially have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of the total period was donated, rather than the observed 12-month frequency. In the assignment step, the donor's proportion of the total period was multiplied by the recipient's maximum possible number of days in the year on which he or she could have used the substance in order to arrive at a 12-month frequency-of-use value for the recipient. Separate assignments were performed within each of the three age groups, subject to the constraints described in the next section. For the 12-month frequency of use, "level of usage" for the State rank groups was defined in terms of the proportion of a given State's residents who had used a given drug in the past month. Assignments were not required for tobacco because the tobacco module did not have 12-month frequency-of-use questions, or for "pills" for the reasons given in the previous section. The final 12-month frequency-of-use imputations were multivariate across drug measures and are further described in **Section 6.3.5**.

6.3.2.6 Constraints on Univariate Predictive Mean Neighborhoods. An obvious logical constraint for 12-month frequency of use was that all donors must be past year users, whether that past year use is reported or (provisionally) imputed. Other logical constraints involved the interview date, month of first use, birthday, recency of use, and 30-day frequency of use. A complete listing of missingness patterns, and the logical constraints associated with those missingness patterns, is given in **Appendix I**. See **Section 6.3.5** for a discussion of the multivariate imputation of recency and frequency of use.

Two likeness constraints used in the assignment of values for 12-month frequency of use were identical to those of recency of use: the three age groups and the State rank groups based on level of past month usage. As with the recency-of-use models, delta was set so that the predicted means of all potential donors were within 5% of the item nonrespondent's predicted mean, where the predicted mean was defined to be the proportion of the year (or maximum period within a year) during which a respondent used a drug. Finally, recipients and donors were required to have the same recency of use (past month vs. past year not past month), whether that recency of use was reported or imputed.[46] 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

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

closest predicted mean was chosen; (2) donors and recipients were no longer required to be from States with similar usage levels; (3) donors and recipients were no longer required to have the same recency of use.

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

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

Even though cocaine and crack are in distinct modules in the CAI questionnaire, a 12-month frequency-of-use model was only fit for cocaine. Donors for crack and cocaine were obtained using a single neighborhood, which was defined in the same manner as for the other drugs.[47] As with recency of use, use or nonuse of crack was considered given (using information from the lifetime imputations). In the same manner as for the drugs where univariate assignments were required, recipients and donors were required to have the same cocaine recency of use, whether that recency of use was reported or imputed. In addition, donors and recipients were also required to have the same crack recency of use if the recipient used crack in the past year.[48] Both of these constraints were applied whether the recipient was missing the 12-month frequency for only cocaine, only crack, or both. Additional logical constraints involved the product of the donated proportion and the recipient's maximum possible number of days used in a year (called the donated 12-month frequency product) for both crack and cocaine. If the 12-month frequencies for both crack and cocaine were missing, this 12-month frequency product for crack could not be greater than that of cocaine. If only the crack 12-month frequency was missing, the donated 12-month frequency product for crack could not be greater than the observed cocaine 12-month frequency; conversely, if only the cocaine 12-month frequency was

---

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

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

missing, the donated 12-month frequency product for cocaine could not be less than the observed crack 12-month frequency. Finally, if the observed 12-month frequency for cocaine was 1, and the 12-month frequency for crack was missing but the respondent was a past year user of crack, naturally the 12-month frequency for crack should be 1.

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

6.3.3 30-Day Frequency of Use

Unlike the recency and 12-month frequency of use, the 30-day frequency of use was not statistically imputed in past NHSDAs; instead, missing values were excluded from subsequent analyses. In the CAI sample of the 1999 NHSDA, however, missing values for 30-day frequency were imputed.

6.3.3.1 Hierarchy of Drugs. Although the imputations were not sequential, provisional recency-of-use indicators for all drugs (except the drug of interest)[49] and the provisional 12-month frequency-of-use variable for the drug of interest (where applicable) served as covariates in the models. Therefore, all recency-of-use imputations and the 12-month frequency-of-use imputation for the drug of interest (where applicable) had to be completed before modeling the 30-day frequency of use.[50] 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 3**).

6.3.3.2 Setup for Model Building and (for Alcohol Only) Hot-Deck Assignment. The file was first subset down to the eligible population: past month users, as defined by the provisional recency variable. Then, item respondents and nonrespondents were defined according to the same criterion that was used for the recency and 12-month frequency

[49]Because the 30-day frequency models were limited to past month users, only one recency category was relevant for the drug of interest. Hence, recency of use for the drug of interest could not be included in the 30-day frequency-of-use model

[50]If the recency of use for a particular drug was not available at the time of the 30-day frequency-of-use imputation for that drug, the lifetime indication of use served as a predictor

imputations. To be an item respondent, the individual had to provide valid responses to all applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of the drug. Predictors for the response propensity models included age; race; gender; census region; an MSA indicator; provisional 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).

6.3.3.3 Model Building. As was apparent from the previous section, only past month users of the drug of interest were used to build the 30-day frequency-of-use model. The (untransformed) response variable of interest in the 30-day frequency-of-use models for most drugs was the proportion of the days in a month (30) on which a respondent used a particular \drug. The range of values for the proportion was from (greater than) 0 to 1. Hence, to model 30-day frequency of use, the following empirical logit transformation was computed for all respondents:

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

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

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

where P_i was defined as the proportion of days in the past year on which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[51] Using the adjusted weights, a linear univariate regression model was then fit for the log-transformed variable Y_i within each age group.

The recency-of-use modeling and 30-day frequency-of-use modeling were performed at almost the same time. Hence, only some of the recency-of-use variables for drugs other than the one of interest were available when the 30-day frequency of use was being modeled for a

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

particular drug. 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 age; age squared; age cubed; gender; race; marital status; employment; educational level[52]; census region; an MSA indicator; recency-of-use values for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; the provisional 12-month frequency of use for the drug of interest (where applicable); and the first- order interactions of age, age squared, age cubed, gender, and race. Because interest was only in the estimation of the predicted mean, and not in the parameter estimates (by themselves) or their standard errors, no model selection was attempted. The predicted 30-day frequencies of use were defined by back-transforming the predicted values from the models. For a complete summary of the 30-day frequency-of-use models, see **Appendix G**.

The predicted mean that comes out of the 30-day frequency-of-use model is a logit of the proportion of the month used. This logit was transformed back into a proportion for use as the variable from which the neighborhoods were created. If the proportion is interpreted as a probability (i.e., the probability of use on a given day in the month conditional on past month use[53]), it could be multiplied by the probability of past month use to make the predicted mean conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it is not known whether they are past month users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

For cigarettes, snuff, and chewing tobacco, the empirical distribution for 30-day frequency of use was in fact a mixture distribution, with a positively skewed distribution from 1 to 29, and a spike at 30. These substances were modeled using two separate models. One was a logistic model for daily use versus nondaily use among past month users. For the nondaily past month users (i.e., those who had used between 1 and 29 days), a model much like the 30-day frequency-of-use models for other substances was used in which the response variable in a linear regression model was a logit of the proportion of the period (30 days) during which a respondent

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

used the substance. The same pool of covariates was used in the logistic model and the regression model with the logit as the response variable. It should be noted that, unlike recency of use, the 30-day frequencies for snuff and chewing tobacco could not be combined into a single value for smokeless tobacco. One could not know if *x* days using snuff overlapped with the *y* days using chewing tobacco. Hence, separate models were fit for snuff and chewing tobacco.

6.3.3.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods. Within a given drug, predicted means from the 30-day frequency-of-use models were computed for both item respondents and item nonrespondents using the parameters from the regression model. The 30-day frequency models were fit after recency of use and 12-month frequency of use. The only drug for which provisional 30-day frequency values were required was alcohol because provisional 30-day frequencies were required to calculate 30-day binge drink provisional values. Neighborhoods were created for each alcohol item \nonrespondent using the UPMN technique described in **Appendix D**. The predicted means used to create the neighborhoods were given by the product of the predicted proportion of the month used (conditioned on past month use) and the probability of past month use given lifetime use (taken from the recency-of-use models).

6.3.3.5 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.

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

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

6.3.4 30-Day Binge Drinking Frequency

In addition to the 30-day frequency of use, an additional frequency variable was defined for alcohol: the number of days in the past month during which the respondent had five or more drinks, or the 30-day binge drinking frequency, also known as DR5DAY. The imputation of the 30-day binge drinking frequency was similar to the imputation of 30-day frequency of alcohol use; however, the 30-day binge drinking frequency model included the provisional alcohol 30-day frequency of use[54] as a covariate. Moreover, the model was built using all past month users of alcohol, whether they were binge drinkers or not. Item respondents for alcohol were defined across recency, 12-month frequency, 30-day frequency, and the 30-day binge drinking frequency measures; therefore, the same weight adjustment was used in the modeling of the 30-day binge drinking frequency as was used for the 30-day frequency model.

The (untransformed) response variable of interest in the 30-day binge drinking frequency models for most drugs was the proportion of the days in a month (30) on which a respondent drank five or more drinks. The range of values for the proportion was from 0 to 1. Hence, to model 30-day frequency of use, the following empirical logit transformation was computed for all respondents:

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

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

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

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

where P_i was defined as the proportion of days in the past month during which respondent i had five or more drinks. The standard logit transformation was not used because it was not defined for daily binge drinkers, nor was it defined for nonbinge drinkers among past month users.[55] Using the adjusted weights, a linear univariate regression model was then fit for the log-transformed variable Y_i within each age group.

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

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

Sections 6.3.1, **6.3.2**, and **6.3.3** summarize how the set of lifetime drug users in the CAI sample of the 1999 NSHDA was separated into item respondents and item nonrespondents for the recency of use, 12-month frequency of use, 30-day frequency of use, and (for alcohol) 30-day binge drinking frequency drug use measures. These sections also summarize model building, computation of predicted means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predicted mean. In most cases, however, these univariate assignments were only provisional. As is indicated in **Exhibit 3**, the final imputed values for these drug use measures were obtained using neighborhoods built upon a vector of predicted means, using the MPMN technique described in **Appendix D**. In a manner consistent with the univariate imputations, the multivariate assignments were done separately within three age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents 26 years of age or older. As indicated in earlier sections, a respondent was eligible to be a donor for a given item nonrespondent if he or she had complete data across the drug use measures for the drug in question and was within the same age group.

The logical constraints required in the univariate imputations discussed in **Sections 6.3.1**, **6.3.2**, and **6.3.3** were also required in the multivariate imputations. In general, the application of these constraints depended on what information was missing in the recency-of-use and frequency-of-use variables. The values that are missing for a given respondent define the "pattern of missingness." For example, one pattern of missingness for marijuana could be as

[55]If the respondent was a daily binge drinker of alcohol, then $\log[(Y+0.5)/(N-Y+0.5] \simeq \log[N+0.5/0.5]$, 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 $\log[(Y+0.5)/(N-Y+0.5] \simeq \log[0.5/(N+0.5)]$. (See Cox and Snell, 1989, for a discussion of the empirical logistic transformation.)

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. 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 J** (see Section J.2).

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

The full predicted mean vector contained several elements for recency of use (different probabilities associated with each of the recency categories), as well as elements for the frequency-of-use variables. Each element in the full vector of predicted means was adjusted so that all elements were conditioned on the same usage status whenever possible. The resulting elements in the predicted mean vector that could potentially result are given in **Exhibit 7**. Note that not all drugs contained all the elements given. **Exhibit 8** shows the full predicted mean vector for each drug. The portion of the full predicted mean vector that was 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 predicted mean vector that were used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given in **Appendix J** (see Section J.3). The Mahalanobis distance was then calculated using only the portion of the predicted mean

Exhibit 7. Elements of Full Predicted Mean Vector

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

[1] Note that the final category for recency (lifetime but not past year, or lifetime but not past 3 years) is not needed in the predicted mean vector because the multinomial probabilities add to 1, and this probability is determined by the other probabilities.
[2] Interpreting the proportion of the year used as a probability of use on a given day in the year assumes that the probability of use on each day in the year is equal. This, of course, is not true. However, the violation of this assumption does not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it does allow the predicted mean to be made conditional on what is known.

Exhibit 8. Full Predicted Mean Vector for CAI 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" 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.

vector that was associated with the given missingness pattern, with elements appropriately adjusted.[56] If no donors were available that had predicted means within a multivariate delta of the recipient's vector of predicted means, the neighborhood was abandoned, and the respondent with the closest Mahalanobis distance was selected as the donor. The procedure is described in detail in **Appendix D**.

6.4 Age at First Use and Related Variables

Unlike the recency and 12-month frequency-of-use variables, age at first drug use was not statistically imputed in past NHSDAs; instead, missing values were excluded from subsequent analyses. As with the 30-day frequency, however, missing age at first use values were imputed for the first time in the CAI sample of the 1999 NHSDA. Also, for the first time in the NHSDA series, 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, missing year and month of first use for less recent initiates (and recent initiates who did not report year and month of first use) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised age at first use, and imputation-revised recency and frequency variables. To have complete date of first use information, day of first use was randomly assigned for all users. The combined data give the respondent's age at first use along with the date of first use. 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.

6.4.1 Age at First Use

The age at first drug use imputations followed the same general procedures as the imputation of other drug use measures. A linear regression model was chosen that was based on a log transformation of the respondent's age at first drug use. UPMNs were formed using the predicted mean from the regression model. Each item nonrespondent's neighborhood was restricted by logical constraints (which cannot be loosened) and likeness constraints (which can be loosened). From these neighborhoods, a final imputation-revised age at first use was created. In addition, a randomly assigned date (i.e., year, month, and day) of first use was constructed that remained consistent with the imputed age at first drug use and other drug use measures.

[56]See **Appendix D** for a definition of Mahalanobis distance.

6.4.1.1 Hierarchy of Drugs. The first step in the imputation of age at first use was to determine the order in which drugs would be modeled. As with the other drug use measures, it was expected that age at first use of other drugs would be strong predictors of age at first use of each drug of interest. Therefore, a hierarchy was chosen in order to get the greatest benefit from using the previously imputed age at first use values as predictors for the drug of interest. The hierarchy for age at first use was identical to the lifetime usage hierarchy given in **Exhibit 4**.

6.4.1.2 Setup for Model Building and Hot-Deck Assignment. As with the imputation of other drug use measures, the file was broken into three age categories for the imputation of age at first use (12 to 17, 18 to 25, and 26 or older), and all subsequent procedures were performed separately within each of these age groups. To impute missing age at first use for each drug, it was necessary to define the eligible population. Using the imputed recency of use, the files were subset down to lifetime users for each drug. If a valid response was provided for the age at first use measure, the person was deemed an item respondent. Before modeling, the respondent weights were adjusted, using a response propensity model, to match the entire population of lifetime users (see **Appendix C** for the more general GEM) and included the following categorical covariates: age, race, gender, census region, MSA, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

6.4.1.3 Sequential Model Building. After the weight adjustment, the following log transformation was calculated for all lifetime drug users:

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

and where i is the drug in question and Y_i is the dependent variable in a weighted linear univariate regression. Variables included in the regression equation were[57] age; age squared; age cubed; State rank (based on the recency variable, see **Section 6.3.1** for details); gender; race/ethnicity; first-order interactions of age, gender, and race/ethnicity; marital status; educational level; employment status[58]; census region; MSA; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin,

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

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

hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; a modified version of the imputed age at first drug use for previously imputed drugs; and modified 12-month and 30-day frequencies for the drug in question. The modified variables for age at first use, 12-month frequency of use (where applicable), and 30-day frequency of use (where applicable) were defined as follows:

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

Naturally, the full model for age at first use did not include the lifetime indicator for the drug in question because the model was built on users of this substance. A summary of the final models can be found in **Appendix G**.

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

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

6.4.1.6 Constraints on Univariate Predictive Mean Neighborhoods. As with all other drug use measures, neighborhoods for age at first use were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, or 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. In fact, recipients and donors were required to be of the same age, if possible. If a donor could not be found of the same age, the constraint eventually reduced to a logical constraint, where the imputed age at first use was less than the recipient's age. A small delta could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. Initially, the relative distance for determining age at first use imputation neighborhoods (delta) was set so that any potential donor's predicted age at first use was within 5% of the recipient's predicted age at first use, and donors were further required to be the same age as the recipient. Another likeness constraint required that if the item nonrespondent had used the drug in the past year, the donor also had to have used it in the past year. Tobacco users had an additional likeness constraint: If the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years. Finally, an attempt was made to require donors and recipients to be from States with similar usage levels, where usage was defined in terms of the prevalence of past month usage of the drug in question.

These likeness constraints were more stringent than those for the other drug use measures. It was often necessary, therefore, to loosen the constraints. The order of loosening constraints follows: (1) remove the State rank group; (2) abandon the neighborhood, and choose the donor with the closest predicted mean; (3) remove the requirement that recipients who were users in the past year (or past 3 years for tobacco) had to have donors who used in the past year (or past 3 years for tobacco); (4) loosen the restriction that donors and recipients had to be the same age, and instead require that the donor's age be greater than or equal to the recipient's age and the donor's age at first use be less than or equal to the recipient's age at first use[59]; and (5) loosen the "same-age" restriction even further, so that the donor's age at first use could be less than or equal to the recipient's age. A summary of the above constraints and the number of respondents who fit into each one is listed for each drug in **Appendix I.**

For drugs with no multivariate assignment, there were several logical constraints. Respondents with an age at first use equal to the recipient's current age were excluded under the following circumstances. First, if the recipient's 12-month frequency was greater than the number of days since his or her last birthday, donors whose age at first use was equal to the recipient's current age were excluded. For example, suppose an item nonrespondent's birthday was on March 1st, and the interview date was June 30th. Then the number of days between the interview date and the respondent's birthday would be 90. If the respondent had a 12-month

---

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

frequency of 100 (either reported or imputed), his or her age at first use could not be his or her current age. In addition, if the respondent's recency of use indicated that he or she did not use in the past month, but the number of days since his or her last birthday was fewer than 30, the recipient's age at first use could not be equal to his or her current age. Finally, if the respondent was not a past month user, but the difference between his or her 12-month frequency and the days since his or her last birthday was fewer than 30, the recipient's age at first use could not be equal to his or her current age. For example, if the recipient respondent's birthday was on March 1st, and the interview was on June 30th, the number of days between the interview date and the respondent's birthday would be 90. If the respondent's 12-month frequency was not a past month user but his or her 12-month frequency was 80, some of those 80 days had to have occurred before his or her birthday, and the respondent's age at first use could not equal his or her current age. Some additional logical constraints were that the donors could not be past year users if the recipient was not a past year user, and, for tobacco, donors could not be users in the past 3 years if the recipient was not a user in the past 3 years. These constraints prevented item nonrespondents from receiving a donated age at first use more recent than the last time they used a substance. Finally, cigarettes had yet another logical constraint: If the recipient was a daily cigarette user and his or her age at first daily use was not missing, the donors were prevented from having an age at first use later than the preexisting age at first daily use.

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

6.4.1.7.1 Smokeless Tobacco (Chewing Tobacco and Snuff). For reasons discussed in **Section 6.2.7.1**, one model for smokeless tobacco was fit rather than individual models for chew and snuff. The nearest neighbor hot-deck neighborhood was then based on the overall smokeless tobacco predicted age at first use. Missing age at first use values for chew and/or snuff were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and if both chew 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, chew donors could not logically be past year chewing tobacco users if recipients were not past year chewing tobacco users. Similar rules applied to snuff (past year and past 3 years) and chew (past 3 years). The likeness constraints were also applied to both chew and snuff separately, but when loosened, they were loosened for chew and snuff simultaneously. Note that, for both chew and snuff, lifetime usage was considered known (employing the lifetime usage imputation), so that there was no question of use versus nonuse of chew or snuff. If age at first use was missing for snuff or chew in the original data, but the respondent was imputed to be a nonuser of snuff or chew in the lifetime imputation, the respondent's age at first snuff use or age at first chew use would be adjusted to reflect the situation. Age at first use for smokeless tobacco was obtained by taking the minimum age at first use from snuff and chew.

6.4.1.7.2 Cocaine and Crack. Even though cocaine and crack are in distinct modules in the CAI questionnaire, an age at first use model was only fit for cocaine. The nearest neighbor hot-deck neighborhood was then based on the overall predicted age at first use for cocaine. Missing age at first use values for cocaine and/or crack were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and if both cocaine and crack were missing, the imputed values came from the same donor. As for the constraints on the neighborhoods, all the constraints listed in the previous section were applied to both cocaine and crack separately. For example, donors for cocaine were logically restricted so that, if the recipient's 12-month cocaine frequency was greater than the number of days since his or her last birthday, donors whose age at first cocaine use was equal to the recipient's age were excluded. The same was true for crack. As a second example, cocaine donors could not logically be past year cocaine users if recipients were not past year cocaine users. Similar rules applied to past year crack use. The likeness constraints were also applied to both cocaine and crack separately, but when loosened, they were loosened for cocaine and crack simultaneously. Note that, for both cocaine and crack, lifetime usage was considered known (employing the lifetime usage imputation), so that there was no question of use versus nonuse of cocaine or crack. If age at first use was missing for crack in the original data, but the respondent was imputed to be a nonuser of crack in the lifetime imputation, the respondent's age at first crack use would be adjusted to reflect the situation.

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

6.4.1.7.3 Stimulants (Methamphetamines and Other Stimulants). As stated in **Section 6.2.7.3**, the stimulants' module included a subgate question referring to methamphetamines, which is of interest in its own right. One model was fit for stimulants' age at first use, from which a single neighborhood was created for both methamphetamines and stimulants as a whole. The nearest neighbor hot-deck neighborhood was then based on the overall stimulants' predicted age at first use. Missing ages at first use for methamphetamines and/or stimulants as a whole were replaced with the values from a donor within this neighborhood. Only missing values were replaced, and if both methamphetamines and stimulants as a whole were missing, the imputed values came from the same donor. As for the constraints on the neighborhoods, the constraints listed in the previous section were all applied to stimulants as a whole. Because no 12-month frequency was available for methamphetamines, however, it was not possible to implement any constraints on methamphetamines involving the 12-month frequency.

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

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

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

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

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

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

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

6.4.1.8.1 Missingness Pattern 1. The first type of missingness pattern occurred when the respondent first starting using the drug 2 years or more before his or her current age. This case is analogous to prior year's data where month and year were not asked in the questionnaire. Below is a brief description of the process involved in obtaining a continuous date of first use in such cases. The imputed YFU, MFU, and DFU were extracted from the continuous date defined below.

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

where

Days between earliest and latest = latest possible date-earliest possible date
Earliest possible date = (integer age at first use*365.25)+birth date
Latest possible date = minimum [(interview date-12 month frequency -1), (earliest date + 365)] *if recency=1*
minimum [(interview date - 29 - 12-month frequency), (earliest date + 365)] *if recency = 2*
minimum [(interview date -1 day - 1 year), (earliest date + 365)] *if recency = 3*
minimum [(interview date - 1 day - 3 years), (earliest date + 365)] *if recency = 4*

6.4.1.8.2 Missingness Pattern 2. The second missingness pattern occurred when the respondent recently initiated use (i.e., within 2 years of his or her current age), and the respondent provided his or her YFU but did not provide an MFU. In such cases, a month and day were randomly assigned that were consistent with both the respondent's frequency/recency and with the age at first use range. The imputed MFU and DFU were derived in the same manner as the date of first use in Pattern 1 with the following changes:

1. If the earliest possible date < YFU, the earliest date = YFU (using January 1^{st} as the earliest month/day).

2. If the latest possible date > YFU, the latest date = YFU (using December 31^{st} as the latest month/day).

6.4.1.8.3 Missingness Pattern 3. Similar to Pattern 2, the third missingness pattern occurred when the respondent recently initiated use (i.e., within 2 years of his or her current age). However, these respondents provided neither an MFU nor a YFU value. In these cases, the year/month/day of first use were randomly assigned from a uniform distribution in a way that was consistent with both the 12-month frequency/recency and age at first use. Again, the imputed YFU, MFU, and DFU were derived in the same manner as described in Pattern 1.

6.4.1.8.4 Missingness Pattern 4. In this case, the respondent provided all the information asked by the questionnaire (i.e., both the month and year of first use). However, to obtain a complete date of first use, a day of first use was also needed. Thus, a day of first use was randomly assigned given the respondent's month and year of first use from a uniform distribution in a way that was consistent with both the 12-month frequency/recency and age at first use. Again, the imputed DFU was derived in the same manner as described in Pattern 1 with the following changes:

1. If the earliest possible date < reported combination of MFU/YFU, the earliest date = MFU/YFU (using 1^{st} day of the month).

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

6.4.1.8.5 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 + 365)] *if recency = 1*
minimum [(interview date - 30), (earliest date + 365)] *if recency = 2*
minimum [(interview date - 1 day - 1 year), (earliest date + 365)] *if recency = 3*
minimum [(interview date - 1 day - 3 years), (earliest date + 365)] *if recency = 4***.**

Another variation occurred with the smokeless tobacco date of first use. In this case, the minimum of the chewing tobacco and snuff date was used to produce the smokeless tobacco date of first use. In addition, the combination drugs (i.e., cocaine and crack, stimulants and methamphetamines) had more constraints placed on their assignment of the dates of first use. Because of the complex relationship between these drugs, the cocaine date of first use was made to be consistent with the crack date of first use and vice versa using both cocaine and crack age of first use data, both recency and frequency data, and any given month/year of first use data for either drug (the same was done for stimulants/methamphetamines).

6.4.2 Age at First Daily Cigarette Use Imputations

In addition to age at first use, the cigarettes' module also included a question asking for the respondent's age at first cigarette daily use, where a daily user was defined as someone who reported having at some time smoked cigarettes every day for a period of at least 30 days. Imputation procedures for age at first cigarette daily use were similar to age at first use, with one key exception: Whereas the age at first use question was asked of all cigarettes 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[60] with a "yes," he or she was considered a daily user. At this stage in the process, there should have been no missing responses to the 30-day frequency question; daily users, based on 30-day frequency, should have been either known (based on a response in the survey) or imputed. However, missing responses for the ever-daily-used question also had to be imputed.

Thus, the age at first daily use imputation involved two parts. First, missing values in the ever-daily-used question (CG15), which asks the respondent if he or she had ever smoked

---

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

everyday for at least 30 days, were imputed.[61] Next, all missing age at first daily use values for eligible daily users were imputed.

6.4.2.1 Setup for Model Building—Ever-Daily-Used Question (CG15). Because age at first daily use was asked of all persons who answered the ever-daily-used question with a "yes," it was necessary to ensure that this question had no missing values. As with all other drug use imputations, the file was broken into three age categories (12 to 17, 18 to 25, and 26 or older), and all subsequent procedures were performed separately within these age groups. To impute for missing values in the ever-daily-used question, it was necessary to define the eligible population: respondents who had an imputation-revised 30-day frequency[62] 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 C** for the more general GEM) and included the following categorical covariates: age, race, gender, census region, MSA, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

6.4.2.2 Model Building—Ever-Daily-Used Question (CG15). After the weights were adjusted, the ever-daily-used question was modeled using weighted logistic regression. Variables included in the initial regression equation were age; age squared; age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of age, gender, and race/ethnicity; marital status; educational level; employment status[63]; census region; MSA; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; a revised 30-day cigarette frequency variable (in the same format as used in the age at first use models, see **Section 6.4.1.3**); and the imputation-revised cigarette age at first use. A summary of the final models can be found in **Appendix G**.

---

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

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

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

6.4.2.3 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Question (CG15). From the final model, a predicted mean of the ever-daily-used question was computed for each eligible respondent. The assignment of imputation-revised ever-daily-used values was conducted using UPMN imputation, as described in **Appendix D**, where the "predicted mean" was the predicted probability of daily use at some point in the respondent's lifetime, given the respondent was a lifetime user but not a current daily user. Again, the procedure defined a "neighborhood" of respondents (i.e., potential donors) by requiring that a respondent's predicted ever-daily-used probability be within a certain relative distance, delta, of the nonrespondent's predicted probability in order to be included in the neigborhood. Delta was set so that donors were required to have a predicted probability within 5% of that of the item nonrespondent.

6.4.2.4 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.

6.4.2.5 Constraints on Univariate Predictive Mean Neighborhoods—Ever-Daily-Used Question (CG15). As with all other drug use measures, neighborhoods for the ever-daily-used question were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, or 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. The likeness constraints were nearly identical to those of age at first use (see **Section 6.4.1.6**). In particular, recipients and donors were required to be of the same age and from States with similar usage levels, if possible. A small delta could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. Initially, the relative distance for determining age at first use imputation neighborhoods (delta) was set so that any potential donor's predicted age at first use was within 5% of the recipient's predicted age at first use, and donors were further required to be the same age as the recipient. The recency likeness constraints that were the same as with age at first use for cigarettes follow: (1) if the item nonrespondent had used in the past year, the donor also had to have used in the past year; and (2) if the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years. Two additional likeness constraints were used as logical constraints when they were applied to age at first use for cigarettes: (1) donors could not be past year users if recipients were not past year users; and (2) donors could not be users in the past 3 years if recipients were not users in the past 3 years.

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

6.4.2.6 Setup for Model Building—Age at First Daily Cigarette Use. After producing an imputation-revised ever-daily-used variable, the next step was the imputation of age at first daily cigarette use values. The eligible population for age at first daily use incorporates all cases deemed to be daily users for at least 30 days at some point in their lifetime. In other words, eligible respondents either had an imputation-revised 30-day cigarette frequency of 30 days or an imputation-revised ever-daily-used value indicating a period in which the respondent smoked everyday for at least 30 days.[64] The file was broken down into three age categories (12 to 17, 18 to 25, and 26 or older), and all subsequent procedures were performed separately within these age groups. If a valid response was provided for the age at first daily use question, the person was deemed an item respondent. Before modeling, the item respondents' weights were adjusted to match the entire eligible population. These adjusted weights were computed using a response propensity model (see **Appendix C** for the more general GEM) and included the following categorical covariates: age, race, gender, census region, MSA, and imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives.

After the weights were adjusted, age at first daily cigarette use was modeled using a weighted linear univariate regression with the dependent variable undergoing the same log transformation as the one defined for the age at first use procedure (see **Section 6.4.1.3**). Variables included in the initial regression equation were age; age squared; age cubed; State rank (based on the recency variable); gender; race/ethnicity; first- and second-order interactions of

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

age, gender, and race/ethnicity; marital status; educational level; employment status[65] ; census region; MSA; imputed recency of use for cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives; modified 30-day cigarette frequency (in the same format as used in the age at first use models); and imputation-revised cigarette age at first use. A summary of the final models can be found in **Appendix G**.

6.4.2.7 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods—Age at First Daily Cigarette Use. From the final model, a predicted mean (based on the *Y* variable) was computed for each eligible daily cigarette user. Then a predicted age at first daily use was derived by back-transforming the predicted mean. The imputation-revised age at first daily use assignment was conducted using UPMN imputation. The procedure defines a "neighborhood" of respondents by 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.

6.4.2.8 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.

6.4.2.9 Constraints on Univariate Predictive Mean Neighborhoods—Age at First Daily Cigarette Use. As with all other drug use measures, neighborhoods for age at first daily use were restricted so that candidate donors and recipients would be within the same age group (12 to 17, 18 to 25, or 26 or older). Models were built separately within these three groups, so this likeness constraint was never loosened. The likeness constraints were nearly identical to those of age at first use (see **Section 6.4.1.6**). In particular, recipients and donors were required to be of the same age and from States with similar usage levels, if possible. A small delta could also be considered a likeness constraint, which could be loosened by enlarging or removing delta. Initially, the relative distance for determining age at first daily use imputation neighborhoods (delta) was set so that any potential donor's predicted age at first daily use was within 5% of the recipient's predicted age at first daily use, and donors were further required to be the same age as the recipient. The recency likeness constraints were the same as with age at first use for cigarettes: (1) if the item nonrespondent had used in the past year, the donor also had to have

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

used in the past year; and (2) if the item nonrespondent had used in the past 3 years, the donor also had to have used in the past 3 years.

The likeness constraints on the donors were loosened in the following order until a neighborhood of at least one donor was achieved: (1) remove the State rank group; (2) abandon the neighborhood, and choose the donor with the closest predicted mean; (3) remove the requirement that recipients who were users in the past year (or past 3 years for tobacco) had to have donors who used in the past year (or past 3 years for tobacco); (4) loosen the restriction that donors and recipients have to be the same age, and instead require that the donor's age be greater than or equal to the recipient's age and the donor's age at first daily use be less than or equal to the recipient's age at first daily use[66]; and (5) loosen the "same-age" restriction even further, so that the donor's age at first daily use could be less than or equal to the recipient's age. A summary of the above constraints, and the number of respondents who fit into each one, is listed for each drug in **Appendix I**.

All the logical constraints applied to cigarettes' age at first use were also applied to age at first daily cigarette use. See **Section 6.4.1.6**, with the words "age at first use" replaced with "age at first daily use." An additional logical constraint was applied specifically to age at first daily cigarette use: If the age at first use for a recipient with a missing age at first daily use was not missing, the donors were prevented from having an age at daily first use earlier than the preexisting age at first use.

6.4.2.10 Date of First Daily Cigarette Use Assignments. After the imputation-revised cigarette age at first daily use was created, all daily cigarette users had a valid age of first daily cigarette use. From this age, a year/month/day of first daily use was assigned. Unlike age at first drug use, the questionnaire did not ask any respondents for their year or month of first daily use of cigarettes. Therefore, the assignment procedure was similar to missing Pattern 1 for age at first drug use (see **Section 6.4.1.8**). Below is a brief description of the process involved in obtaining a continuous date of first daily cigarette use.

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

where

Days between earliest and latest = latest possible date - earliest possible date
Earliest possible date = (integer age at first use * 365.25) + birth date

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

Latest possible date = minimum [(interview date - 30 day frequency + 1), (earliest date + 365)] *if recency = 1*
minimum [(interview date - 30), (earliest date + 365)] *if recency = 2*
minimum [(interview date - 1 day - 1 year), (earliest date + 365)] *if recency = 3*
minimum [(interview date - 1 day - 3 years), (earliest date + 365)] *if recency = 4*

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

7. CAI Income and Insurance Imputations

This chapter summarizes the techniques used to edit and impute missing values in the income and insurance variables. As with the drug imputations discussed in **Chapter 6**, imputations were accomplished using the predictive mean neighborhood (PMN) technique described in **Appendix D**. However, whereas the editing process for the drug imputations were described in another document (Kroutil, 2001a), the editing procedures implemented on the income and insurance variables are described in the following sections.

7.1 Health Insurance

7.1.1 Edited Insurance Variables

Exhibit 9 summarizes the relationship between a sample of health insurance questionnaire variables and their edited counterparts. The edited variables have the same values as the questionnaire variables, except that missing values are replaced by standard NHSDA missing value codes.

Exhibit 9. Mapping of Questionnaire Health Insurance Variables to Edited Counterparts

Variable	Question	Edited Counterpart
QHI01	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)
QHI02	Is the respondent covered by Medicaid or Medical Assistance?	MEDICAID (1 = yes, 2 = no)
QHI03	Is the respondent covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)
QHI04	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)

These four questionnaire variables were used to create two overall insurance variables: INSUR (respondent has health insurance) and PINSUR (respondent has private health insurance). INSUR was coded as "yes" if any one of the four variables listed in Exhibit 9 were coded as "yes," and INSUR was coded as "no" if all four variables were coded as "no." Missing data in PRVHLTIN were coded using the standard NHSDA missing data codes for "don't know," refused, and blank, whereas missing data in PINSUR were all coded as "98." Otherwise, PINSUR and PRVHLTIN are equivalent. PINSUR was created to maintain consistency with previous NHSDAs, when 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; hence, respondents with private health insurance are a subset of the respondents who have health insurance.

7.1.2 Imputed Health Insurance Variables

7.1.2.1 Hierarchy (Modeling Order) of Health Insurance Variables. A multivariate imputation 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. It was therefore not possible to use INSUR as a covariate in the PINSUR model, or vice versa. As a result, the models for the two variables could be run simultaneously.

7.1.2.2 Setup for Model Building and Hot-Deck Assignment. The next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for both health insurance variables were conducted separately within the four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. Even though both edited insurance variables were part of the same multivariate set, respondents and nonrespondents were determined separately for the two variables. Response propensity adjustments to the weights, to make the item respondent weights representative of the entire sample, were also implemented separately for the two variables within each age group. The item response propensity model is a special case of the generalized exponential model (GEM), which is described in greater detail in **Appendix C**. The variables included in the model predicting the probability of item nonresponse were the same as those included in the main model, which is discussed in the next section.

7.1.2.3 Sequential Model Building. The probability that the respondent had health insurance and the probability that the respondent had private health insurance were both modeled for item respondents, within each age group, using the nonresponse adjusted weights. The parameters for the models were estimated using logistic regression. The predictors considered in each model included continuous age, race/ethnicity, age squared, gender, population density, percentage of housing in segment that is owner-occupied, percentage concentration of Hispanics in segment, percentage concentration of blacks in segment, household size, and one-way interactions of age, age squared, race/ethnicity, and gender. For the three older age groups, the additional predictors of marital status, educational level, and employment status were also considered in each model.

7.1.2.4 Computation of Predicted Means. Using the parameter estimates from the probability of lifetime use model for a given drug, predicted probabilities of use were computed for both item respondents and nonrespondents. Because neither variable could be used as a covariate in the model for the other variable, no provisional values were required.

7.1.2.5 Multivariate Imputation of Health Insurance and Private Health Insurance. The final imputed values for health insurance and private health insurance were obtained using neighborhoods built upon a vector of predicted means (see the MPMN technique description in **Appendix D**). As with the models, 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. A respondent was eligible to be a donor for a given item nonrespondent if he or she had complete data across both health insurance variables and was within the same age group. The set of potential donors was then further restricted to be of 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. In either case, the pool of donors selected were those with the 30 smallest Mahalanobis distances.[67] (If 30 donors did not meet the age constraint but some number of donors fewer than 30 did meet the constraint, the neighborhood consisted of those donors who met the constraint.) The patterns of missingness for each drug, the logical constraints imposed on the set of donors, and the frequency of occurrence of each missingness pattern are given in **Appendix J** (Section J.2). The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix I**.

The full predicted mean vector contained elements for overall health insurance and for private health insurance. The portion of the full predicted mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If one of the two variables was not missing, the predictive mean vector used to determine the neighborhood was limited to the predicted mean associated with the missing variable. The portions of the full predicted mean vector that were used to create the MPMN neighborhoods for each missingness pattern, with accompanying adjustments, are given in **Appendix I (Section I.3)**. The Mahalanobis distance was then calculated using only the portion of the predicted mean vector that was associated with the given missingness pattern. The

[67]Mahalanobis distance is defined in **Appendix D**.

set of donors was then restricted to a neighborhood of 30 or smaller by sorting by the Mahalanobis distances and selecting the donors with the 30 smallest distances, as described in **Appendix D**.

7.2 Income

The imputation of income was separated into two phases. The first phase, the "binary variable phase," involved the imputation of all the binary income variables, as well as the number of months on welfare. This included the "yes-no" questions about the sources of income for the respondent and for the respondent's family living in the respondent's household, the number of months on welfare question (the only nonbinary variable in the binary variable phase), and a "yes-no" question regarding whether the respondent's income or the respondent's family income (in the household) was $20,000 or more (including income from the sources referred to in the previous questions). The correspondence between these questionnaire items and the edited variables is given in **Exhibit 10**. The second phase, the "specific category phase," consisted of imputing more specific income categories for the respondent and the respondent's family in the household.

7.2.1 Edited Income Variables: Binary Variable Phase

7.2.1.1 Source of Income Variables. Most of the variables measuring the source of income consisted of two parts: 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.[68] From these two parts, three edited income source variables were created: 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.

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

Exhibit 10. Mapping of Questionnaire Income Variables to Edited Counterparts

Source of Income/Binary Total Income Questions				
Variable Description	**Raw Questions**	**Edited Personal Income**	**Edited Other Family Income**	**Edited Total Family Income**
Social Security	QI01, QI02	PSOC	OFMSOC	FAMSOC
Supplemental Security	QI03, QI04A, QI04B	PSSI	OFMSSI	FAMSSI
Food Stamps	QI05A, QI05B	-----*	-----*	FSTAMP
Welfare Payments	QI06, QI07A, QI07B	PPMT	OFMPMT	FAMPMT
Other Welfare Services	QI08, QI09A, QI09B	PSVC	OFMSVC	FAMSVC
# Welfare Months	QI10A, QI10B	-----*	-----*	WELMOS
Investment Income	QI11, QI12A, QI12B	PINT	OFMINT	FAMINT
Child Support	QI13, QI14A, QI14B	PCHD	OFMCHD	FAMCHD
Wages	QI15, QI16A, QI16B	PWAG	OFMWAG	FAMWAG
Other Income	QI17, QI18A, QI18B	POTH	OFMOTH	FAMOTH
Total Income	QI20, QI22	PINC1	-----*	FINC1
Total Income Specific Categories	QI21A, QI21B, QI23A, QI23B	PINC2	-----*	FINC2

*Edited variables are not generated.

Every respondent was eligible to answer the personal source of income questions. Hence, the raw and edited personal source of income variables are equivalent. The other-family-member income questions required more editing. As stated previously, if the respondent answered "yes" to the personal question or did not have any family members in the household, the other-family-member question should have been skipped and was coded as a legitimate skip.[69] 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.

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

Food stamps' information was collected using one question (QI05A/QI05B) that applied to the respondent's entire family. The question concerning months on welfare (QI10A/QI10B) was only asked for respondents who answered "yes" to either the welfare payments' (personal, QI06, or other family, QI07A/QI07B) or other welfare services' (personal, QI08, or other family, QI09A/QI09B) source of income questions.

7.2.1.2 Personal and Family Total Income Variables. In addition to the source of income variables, the binary variable phase also included a pair of binary variables regarding whether the respondent's personal total income or the respondent's family's total income was \$20,000 or more. For this pair of questions, the second question in the pair applied to the entire family. As with the source of income variables, the raw and edited personal total income variables were equivalent. The second question in the pair asked about total family total income, but was skipped if the respondent had no other family members in the household. The edited variable was created by assigning legitimate skips in those cases. A third binary family total income variable was created and was equal to the response to the second question in the pair if other family members were present in the household. Conversely, if no other family members were present, it was equal to the response to the first question in the pair. Finally, if the total personal income response indicated an income of \$20,000 or more, but the total family income response was less than \$20,000, the values for all three variables were set to missing and later imputed.[70]

7.2.2 Imputed Income Variables: Binary Variable Phase

7.2.2.1 Hierarchy (Modeling Sequence) of 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 (i.e., the "income hierarchy" discussed in detail in **Appendix D**). For a model predicting whether a respondent had a given source of income, it was expected that other sources of income would be useful covariates. Following a provisional imputation of missing income values in the binary variable phase, the indicators earlier in the sequence were used as covariates for income models later in the sequence. The resulting values were only provisional 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

[70] An error in the 1999 instrument allowed interviewers to enter specific income categories that differed from a respondent's original entry in the binary total income question. For example, it was possible for an interviewer to state that the respondent's personal income was less than \$20,000, but the specific income category he or she entered might have been between \$50,000 and \$74,999 (even though the screen he or she saw indicated categories under \$20,000). In these instances, the specific category was believed, and the original entry in the binary total income question was overwritten.

the binary variable phase. The order in which the income indicators were imputed is shown in **Exhibit 11**.

7.2.2.2 Setup for Model Building. Once the hierarchy of income variables in the binary variable phase was established, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all income indicators were conducted separately within the four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. For an individual to be considered an item respondent for income variables in the binary variable phase, he or she must have complete data for all of the questions included in this phase: social security, supplemental social security, welfare payments and services, investments, child support, wages, other sources of income, food stamps, months on welfare, and total family income (less than $20,000 vs. $20,000 or more). Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. Note that 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 C**. The variables included in the model predicting the probability of item nonresponse were the same as those included in the main model, which is discussed in the next section.

7.2.2.3 Sequential Model Building. Starting with social security, the probability that a family received income from a given source was modeled for item respondents, within each age group, using the nonresponse adjusted weights in a weighted logistic regression model. 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 11.** The pool of covariates that were considered for each model included continuous age; age squared; gender; race/ethnicity; provisional income indicators earlier in the sequence; region; population density; percent Hispanic categories in segment; percent black categories in segment; percent owner-occupied households in segment; 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; a three-level State rank variable; and first-order interactions of age, age squared, race/ethnicity, and gender. For the three older age groups, three other covariates were also considered: marital status, education status, and employment status. The State rank groups were defined in terms of the proportion of a given State's residents whose income was greater than or equal to $20,000. This pool was reduced for

Exhibit 11. 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
Welfare Months	WELMOS
Total Family Income[1]	FINC1

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

each variable using backward elimination; the final sets of covariates used for each variable are given in **Appendix G**.

The same pool of covariates was 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 logit: $Y = \text{logit}(p)$, where p = number of months on welfare ÷ 12. The binary total family income variable was modeled using weighted logistic regression. For a complete summary of the income imputation models, see **Appendix G**.

7.2.2.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods. Following the modeling of each income variable in the binary variable phase, missing values were replaced by provisional imputed values. This was necessary so that these variables could be used as covariates in subsequent models. Although no provisional imputed values were used to build the models, predicted means needed to be calculated for all respondents, including item nonrespondents, using the parameter estimates from the models. This sometimes required the use of the provisional values for the covariates. The predicted probabilities from these models were used to assign provisional values using the UPMN imputation method described in **Appendix D**.

7.2.2.5 Assignment of Provisional Imputed Values. Separate assignments of provisional values were performed within each of the four age groups for all income variables. The final income imputations were multivariate across all the variables in the binary variable phase (i.e., the source of income, months on welfare, and the total income variables). The multivariate imputation process is further described in **Section 7.2.2.8**.

7.2.2.6 Constraints on Univariate Predictive Mean Neighborhoods. After predicted values from the model had been determined, a univariate imputation was implemented on each variable within each age group. If the respondent was missing the value for a particular income variable, a neighborhood was determined by choosing respondents with predicted values "close to" (within 10% of) the recipient's predicted value.[71] If fewer than 30 donors had predicted values within 10% of the recipient's predicted value, any neighborhood of size greater than zero was allowed. If there were still insufficient donors, the predicted means of the donors were allowed to be within 20% of the recipient's predicted value. Donors also were required to have the same value for the family skip variable (IRFAMSKP) as the recipient. If one of the personal or other-family-member variable values was nonmissing, donors and recipients were required to have the same value for the nonmissing variable.

7.2.2.7 Multivariate Assignments. The predicted means were calculated with edited family income variables (see **Exhibit 11**) as the response variables. For each variable, neighborhoods were created using scalar-predicted means from the appropriate model. A univariate methodology, in terms of these scalar-predicted means, was therefore used to determine the neighborhood. In most cases, three edited variables were associated with each predicted mean, so that missing values for three variables required assignment of imputed values. Hence, even when determining the provisional imputed values using the univariate procedure, the assignment of imputed values was multivariate for all but two of the binary phase variables. The variables associated with each of the models are given in **Exhibit 12.**

7.2.2.8 Multivariate Imputation. **Sections 7.2.2.1** through **7.2.2.7** summarize how the set of income variables in the CAI sample of the 1999 NHSDA were separated into item respondents and item nonrespondents. These sections also summarize model building, computation of predicted means and delta neighborhoods, and the assignment of imputed values for these measures using a univariate predicted mean. In most cases, however, these univariate

[71]Subsequent theoretical considerations support the notion that the predicted mean of the donor and recipient must be as close as possible (within 5%), whereby if no donors were available within 5% of the recipient's predicted mean, the closest donor was chosen and a neighborhood was not used. The 1999 drug use imputations were redone using this criterion.

assignments were only provisional. The final imputed values for these drug use measures were obtained using neighborhoods built on a vector of predicted means using the MPMN technique described in **Appendix D**. In a manner 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.

Exhibit 12. 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, IRFAMINC

The source-of-income variables, a single months-on-welfare variable, and the binary total income variables are outlined in **Exhibit 10**. The collective distance between these variables' conditional predictive means for a given incomplete data respondent and the complete data respondents was determined using a Mahalanobis distance within each age group. The donors were usually restricted to have an age the same as the recipient, or if that constraint was too restrictive, an age within 5 years of the recipient was used. Donors also were required to have the same value for the family skip variable (IRFAMSKP) as the recipient. Of the variables outlined in **Exhibit 10**, 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 high-income respondents, 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 (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 the recipient's nonmissing information was the same as the donors information for those variables. If one of the pair of income variables (personal and other-family-member variables) 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) loosen the constraint requiring donors to be of the same age as the recipient to donors having ages within 5 years of the recipient; (2) loosen the constraint that incorporated the association between the welfare, food stamps, and income payment questions; and (3) loosen the requirement that the imputation be multivariate. If a respondent was missing the months-on-welfare question, but was not missing one of the feeders to this question (the questions involving welfare payments or welfare services), the donor and recipient were required to have the same values for the nonmissing feeder question variables. This constraint was never loosened. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix I**.

The 30 complete data respondents with the closest Mahalanobis distance to the recipient constituted the multivariate neighborhood, from which a single donor was randomly drawn. 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 family welfare question, only the donor's family welfare response was given to the recipient. The Mahalanobis distance only included predicted values for missing values. This required the determination of missingness patterns, so that the appropriate Mahalanobis distance could be calculated.

7.2.3 Edited Income Variables: Specific Category Phase

As part of the second phase of the income questions, respondents were asked to identify specific categories of income within the two general categories previously selected (less than $20,000 or $20,000 or more) both for themselves and for their families. For respondents who answered the binary total income question as less than $20,000, they were asked to enter a specific category of income within increments of $1,000 (between $0 and $999, between $1,000 and $1,999, etc.). Conversely, respondents who answered the binary total income question as $20,000 or more were asked to enter a specific category of income within increments of $5,000 up to $50,000 (between $20,000 and $24,999, between $25,000 and $29,999, etc.), or between $50,000 and $74,999, or more than $75,000.[72]

---

[72]An error in the 1999 instrument allowed interviewers to enter specific income categories for respondents that differed from their original entry in the binary total income question. For example, it was possible for an interviewer to state that the respondent's personal income was less than $20,000, but the specific income category he or she entered might be between $50,000 and $74,999 (even though the screen he or she saw indicated categories under $20,000). In these instances, the specific category was believed, and the original entry in the binary total

As with the binary total income questions, the specific category questions were asked in a pair, the first for the individual respondent and the second for the entire family. As with the other variables that followed this 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.[73] The edited variable was created by assigning legitimate skips in those cases. A third specific category family total income variable was created that was equal to the response to the second question in the pair if other family members were present in the household. Conversely, if no other family members were present, it was equal to the response to the first question in the pair. Finally, if the binary total income responses were set to bad data, the specific category responses were also set to bad data.

7.2.4 Imputed Income Variables: Specific Category Phase

7.2.4.1 Hierarchy of Income Variables. Three income variables resulted from editing the questions in the income specific category phase (see **Exhibit 10**). These three variables were all considered together using a failure time model. Because only one model was fit, no hierarchy was required.

7.2.4.2 Setup for Model Building. As with the variables in the binary variable phase, the imputations were conducted separately within the four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. For an individual to be considered an item respondent for income variables in the specific category phase, he or she must have had complete data for both questions in this phase. Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample, and the appropriately adjusted weights were used in the models. The item response propensity model is a special case of the generalized exponential model (GEM), which is described in greater detail in **Appendix C**. 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.

---

income question was overwritten.

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

7.2.4.3 Sequential Model Building. The specific categories of income were modeled using the LIFEREG procedure in SAS. This procedure was used for regression modeling of continuous non-negative random variables, such as survival times and income, fitting models that are sometimes referred to as "failure time models." The model assumed for the response variable y (income) is

$$y = X\beta + \sigma\epsilon$$

where X is the matrix of covariates, β is the parameter vector, σ is a scale parameter, and ε is a vector of error terms, which are assumed to come from some known distribution, such as the logarithm of a three-parameter generalized gamma model, or one of its more common two-parameter special cases (gamma, Weibull, lognormal, or log-logistic). Although the underlying random variable y is assumed to be continuous, the LIFEREG procedure allows the variable to be reported in interval categories, such as the NHSDA income intervals. The contribution of an individual with covariates X to the overall likelihood is just the probability mass assigned by the model to the interval $(l, u]$ containing the actual (continuous) income for that individual. This contribution has the form $F(u|X,\beta,\sigma) - F(l|X,\beta,\sigma)$, where F is the cumulative distribution function of the assumed type. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.

LIFEREG allows several choices for the functional form of the parametric model that correspond to the error distribution discussed earlier, including the two-parameter log-logistic, lognormal, gamma, Weibull, and the three-parameter generalized gamma. Each of these models was fit to each of the four age groups/income datasets, and the log-logistic and gamma distributions provided a better overall fit, as measured by the likelihood, than the other models. When considering data from the 2000 questionnaire, which was also available for comparison, the gamma distribution appeared to fit the data the best. Because the three-parameter generalized gamma did not improve significantly on its two-parameter special cases using likelihood ratio tests, it was decided to use a two-parameter model.

The covariates considered in the model included the same covariates that were used in the binary variable phase: continuous age, age squared, gender, race/ethnicity, all imputation-revised income indicators considered in the binary variable phase, region, population density, percent Hispanic categories in segment, percent black categories in segment, percent owner-occupied households in segment, 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, a three-level State rank variable, and first-order interactions of age, age squared, race/ethnicity, and gender. The State rank groups were defined in terms of the proportion of a given State's residents whose income was greater than or equal to \$20,000.

7.2.4.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods. The "predicted value" $X\beta$ provided by the failure time model was a monotonic function of the conditional mean of the modeled income distribution at a given individual set of values of the regressor variables. (Specifically, $X\beta$ was a translation of the estimated mean of log income.) These values were computed for both item respondents and item nonrespondents using the parameters from the failure time model. They were used to assign imputed values using the UPMN imputation method described in **Appendix D**.

7.2.4.5 Assignment of Imputed Values. Separate assignments of imputed values were performed within each of the four age groups for all specific category income variables. Only missing values were replaced by imputed values using the same donor for all three variables.

7.2.4.6 Constraints on Univariate Predictive Mean Neighborhoods. Donors and recipients were required to have the same values for both the binary income variable and the indicator of whether other family members were in the household (IRFAMSKP). In addition, if either of the personal income or family income specific category responses were nonmissing, donors and recipients were required to have the same values for the nonmissing variable. Finally, donors were required to have predicted values "close to" (within 10% of) the recipient's predicted value.[74] 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. If still no donors could be found, the definition of "close" was loosened to 20%. Finally, if no donors were available, the neighborhood was abandoned, and the donor with the closest predicted mean to the recipient was chosen, subject to the logical constraints described above. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix I**.

7.2.4.7 Multivariate Assignments. The predicted means were calculated using the edited (specific category) family income variables (see **Exhibit 11**) as the response variables.

---

[74]Subsequent theoretical considerations supported the notion that the predicted mean of the donor and recipient had to be as close as possible (within 5%), whereby if no donors were available within 5% of the recipient's predicted mean, the closest donor was chosen and a neighborhood was not used. The 1999 drug use imputations were redone using this criterion.

For each family income variable, neighborhoods were created using scalar-predicted means from the appropriate model. The methodology for determining the neighborhood was therefore univariate in terms of these scalar-predicted means. Three edited variables were associated with each predicted mean, so that missing values for three variables required assignment of imputed values. Hence, even when determining the provisional imputed values using the univariate procedure, the assignment of imputed values was multivariate for all but two of the binary phase variables. The imputation-revised variable for the personal income variable is called IRPINC2, the family income variable with legitimate skips is called IRFINC2, and the family income variable without legitimate skips is called IRFAMIN2.

This page intentionally left blank.

8. Household Composition (Roster) Editing and Imputations

This chapter summarizes the techniques used to edit inconsistent values in the household roster and the techniques used to create and impute missing values in the roster-derived household composition variables. As with the drug imputations discussed in a previous chapter (**Chapter 6**), imputations were accomplished using the predictive mean neighborhood (PMN) technique described in **Appendix D**. However, whereas the editing process for the drug imputations are described elsewhere (see Kroutil, 2001a), the editing procedures implemented on the household roster and the procedures to create the roster-derived household composition variables are summarized in the following sections.

8.1 Household Roster Edits

8.1.1 Description of Household Composition (Roster) Section of Questionnaire

The introductory question to the questionnaire household roster (QD49) asked the respondent (interviewer administered) for information regarding the number of people living in his or her household, where allowable entries ranged from 1 to 25. If either the interviewer indicated that the respondent lived alone or the question was unanswered, the household composition (roster) section was skipped. However, if the interviewer indicated a household size greater than 1, the interviewer was then prompted to ask the respondent questions about the age, gender, and relationship to the respondent of every member of the household, starting with the household's oldest member, and including the respondent. 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 QD49. An example of such a grid when QD49 = 4 is given in **Exhibit 13**. The relationship codes are given in **Exhibit 14**. Also given in **Exhibit 14** are details corresponding to certain relationship codes.

8.1.2 Preliminary Roster Edits

To facilitate processing of the roster variables, a "roster-level" file was created in which the number of records per respondent is given by the household size in QD49. If the respondent broke off the interview after the household size question, or in the middle of the roster questions, "dummy" records were created that corresponded to the missing household members.

Exhibit 13. Household Composition (Roster) Grid Example, QD49 = 4

Person #	Relationship to Respondent	Age in Years
1	Self	44
2	Husband	42
3	Son	16
4	Boarder/Roomer	16

Exhibit 14. Household Composition (Roster) Relationship Codes

Relationship Code #	Relationship to Respondent	Details About Relationship
1	Self	
2	Parent	Biological, Step, Adoptive, or Foster
3	Child	Biological, Step, Adoptive, or Foster
4	Sibling	Full, Half, Step, Adoptive, or Foster
5	Spouse	
6	Living Together As Though Married	
7	Housemate or Roommate	
8	Child-in-Law	
9	Grandchild	
10	Parent-in-Law	
11	Grandparent	
12	Boarder or Roomer	
13	Other Relative	
14	Other Nonrelative	

8.1.3 Roster Edits Involving the Self

If only one roster member was identified as self, where the age of the roster member was within 1 year of the questionnaire-edited age[75] (AGE, defined in **Chapter 4**), and the gender for self matched IRSEX (also defined in Chapter 4), the roster entry for self was assumed to be correct. If this was the case, the roster age was set to AGE, and no further action was required for the self record. However, if this was not the case, it was necessary to either identify a self among the roster records for that respondent, or add another record that could be identified as the self. There were three ways in which an interviewer could enter incorrect information for the self in the household roster: (1) no self in roster, (2) multiple selves in roster, or (3) the roster age for self differed from AGE by more than 1 year, or the gender for self in the roster did not match IRSEX. Each of these self edits is discussed in turn below.

8.1.3.1 Edits for No Self in Roster. If the interviewer did not identify a self in the roster, it was necessary to try to find a self among the roster members corresponding to the respondent in question. A roster member was selected as the self under one of two possible circumstances: (1) the roster member's age, gender, and relationship data were missing, or (2) the roster member was of the respondent's gender, and was within 1 year of the respondent in age, and had a relationship code that was impossible. Only one roster member had a relationship code changed to self. However, it was possible to have (a) more than one roster member with missing information; (b) more than one roster member with a gender that matched the respondent's, an age within 1 year of the respondent, and an impossible relationship code; and (c) any combination of (a) and (b). For the situation described by (a), one of the roster members with missing information was chosen, where the relationship code was set to self, the roster age set to AGE, and the roster gender set to IRSEX. The remaining roster members with missing information were left alone. For the situations described by (b) or (c), it was necessary to select the one among these roster members that would be assigned to the self. Among all the roster-level records with impossible relationship codes that could possibly be reassigned as self, the self code was assigned to the roster member in the following priority order (each of the listed relationships were considered impossible since the ages of the roster member and the respondent differ by a year or less):

1. The roster member was reported as the respondent's biological, adoptive, or foster parent;

---

[75]Because the interview might begin and end on different dates, a respondents age might change between these two dates. As a result of this, a 1-year difference was allowed.

2. The roster member was younger than 15 years old and was reported as the respondent's stepparent;

3. The roster member was reported as the respondent's biological, adoptive, or foster child;

4. The roster member was reported as the respondent's stepchild, but the respondent was younger than 15;

5. The roster member was reported as the respondent's legal spouse, grandchild, or grandparent; or

6. The roster member's relationship, age, and gender data were missing.

If no roster member met the above criteria, it was assumed that the respondent did not consider himself or herself when counting the number of people in his or her household. The value of QD49 was assumed to be wrong (one fewer than necessary), and a record was added with a relationship code of self, a roster age equal to AGE, and a roster gender equal to IRSEX.

8.1.3.2 Edits for Multiple Selves in Roster. If multiple selves were identified in the roster, an attempt was made to identify the correct self among all roster members with a self relationship code. If one or more of the roster members with the self code had a roster age that matched the edited questionnaire age (AGE) exactly, and roster gender matched IRSEX, the first among these roster members was selected as the true self. If no exact match was available, but one or more of the roster members with the self code had a roster age that differed from AGE by a year, with an exact match on IRSEX, the first among these roster members was selected as the true self. Finally, if none of the roster members with the self relationship code had an age-gender approximate match (age within 1 year) with AGE and IRSEX, QD49 was assumed to be wrong (one fewer than necessary), and a record was added with the relationship code of self, a roster age equal to AGE, and a roster gender equal to IRSEX.

8.1.3.3 Edits for Cases When the Assigned Self Did Not Have Appropriate Age or Gender. Although the interviewer might have identified a single roster member as the self, it was possible that the identification was incorrect and that the self may actually have corresponded to a different roster member. Perhaps the interviewer may have applied the wrong relationship codes to the roster members using a household member other than the respondent as the reference point. Using the example given in **Exhibit 13**, if the respondent's son was used as the reference point, the relationship for the respondent became "mother" instead of "self" and the husband became "father." Under these circumstances, the self code was set to missing, and the respondent's roster entries became a no-self household. The procedures for finding the roster member who was the self was then equivalent to the no-self case outlined in **Section 8.1.3.1**.

8.1.4 Roster Edits for Other Household Members

Relationship codes were set to missing if the relationship of the roster member was impossible based on age and gender, and a self code was not assigned. The following relationships were considered impossible:

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

2. The roster member was reported as the respondent's biological parent, but was less than 12 years older than the respondent.

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

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

5. The roster member was reported as the respondent's biological, adoptive, or foster child, but was the same age as or older than the respondent.

6. The roster member was reported as the respondent's biological child, but was less than 12 years younger than the respondent.

In addition, if a roster-level record was listed as a potential self using the edits described in **Section 8.1.3.1**, but was not assigned a self code because another roster-level record was assigned that code, the relationship code was set to missing. Finally, 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.[76]

---

[76]In 2000, edits were added that check whether the grandchild-grandparent relationship codes made sense with respect to the roster and respondent ages.

8.2 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 15.** If any information in the roster was missing, the roster-derived variable was set to missing. However, if some of the roster records for a respondent's household had missing data, roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could be imputed. Details on the imputation of the household roster-derived variables are given in **Section 8.3**.

The respondent's household size was assumed to equal the total number of rostered people in the household, TOTPEOP, as shown in **Exhibit 15**. The value of TOTPEOP was expected to equal to QD49 in most cases. However, in some cases the assigned self did not match, even approximately, the respondent's age or gender, or no self was assigned and no other roster members matched the respondent's age and gender. In these cases, an extra roster member was added to correspond to the respondent (the self), so that the value of TOTPEOP was one greater than QD49. In some cases, the respondent did not enter a value for QD49, 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 would be missing, regardless of whether some of the roster members were eligible to be part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in **Section 8.3.6**. However, if the roster member was missing a relationship code, but not an age, that roster member was still eligible to be counted in these variables.

FAMSKIP was an indicator of whether the respondent's household contained other family members. It was created based on the relationship codes of the roster members. If one or more of the roster members had a missing relationship code, and no other family members were in the respondent's household, the value of FAMSKIP would be set to missing. However, if one of the nonmissing roster member's relationship codes indicated that the household contained one of the

Exhibit 15. 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 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	IMOTHER
Indicator of presence of father in household	IFATHER

respondent's family members, the value of FAMSKIP would not be missing even if other roster members had missing relationship codes.[77]

Nine other roster-derived variables were created that used both the age and relationship codes of the roster members. All of the roster-derived variables and their definitions are summarized in Exhibit 15. Each of these variables was missing if the age or relationship codes for at least one roster member in a respondent's household was missing.

[77]If the roster edits removed all family relationship codes from the household roster, it was possible that a respondent who originally had family members would have received a missing FAMSKIP value, which could have been imputed to indicate no family members. In 2001, a change will be implemented whereby the value of FAMSKIP will be determined by the original relationship codes, even though some of those codes may be impossible. The assumption here was that even though the exact family relationship may have been wrong, the roster member probably was still a family member.

8.3 Imputation of Household Roster-Derived Variables

Although nine roster-derived variables were created from the edited roster, the missing values were imputed for only four of these variables: TOTPEOP, KID17, HH65, and FAMSKIP. The missing values for these variables were imputed using the UPMN technique described in **Appendix D**.

8.3.1 Hierarchy of Household Roster-Derived Variables

After editing the roster variables, the next step in the imputation of household roster-derived variables was to determine the order in which the variables would be modeled. Each roster-derived variable was expected to be strongly related to the other three roster-derived variables. Hence, it was importaner-derived variable was expected to be strongly related to the other three roster-derived variables. Hence, it was important to have a hierarchy so that variables early in the sequence could be used as covariates for subsequent variables. The order in which the roster variables

were imputed is shown in **Exhibit 16**.

Exhibit 16. Household Roster-Derived Variables (in Order of Imputation)

Roster Variable	Edited Variable	Imputed Variable
Total number of rostered people	TOTPEOP	IRHHSIZE
Total number of kids under age 18	KID17	IRKID17
Total number of people aged 65 or older	HH65	IRHH65
Indicator of whether the respondent has family members in household	FAMSKIP	IRFAMSKP

8.3.2 Setup for Model Building

Once the hierarchy of the roster-derived variables was established, the next step was to define respondents, nonrespondents, and the item response mechanism. Imputations for all roster-derived variables were conducted separately within the four age groups: 12 to 17 year olds, 18 to 25 year olds, 26 to 64 year olds, and respondents 65 years of age or older. Response propensity adjustments were then computed for each age group in order to make the item respondent weights representative of the entire sample. Item respondents were not defined across all roster categories; hence, this adjustment was computed separately for each age group and for each variable. The covariates in the response propensity models were the same covariates as those considered for the main model presented in the next section. The item response propensity model is described in greater detail in **Appendix C**.

8.3.3 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 weighted Poisson regression[78]. The binary variable FAMSKIP was modeled using a weighted logistic regression. The covariates considered in each model included continuous age; age squared; race/ethnicity; gender; first- order interactions of age, age squared, race/ethnicity, and gender; region; population density; percent Hispanic households in segment; percent owner-occupied households in segment; number of people in the household eligible for interviewing (from the pre-interview screener); and roster-derived variables earlier in the hierarchy. For all age groups except the 12 to 17 year olds, marital status, education status, and employment status were also included as covariates.

8.3.4 Computation of Predicted Means and Univariate Predictive Mean Neighborhoods

From the final models, a predicted mean was computed for every respondent, from which neighborhoods were determined following the UPMN procedure outlined in **Appendix D**.

8.3.5 Assignment of Imputed Values

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

8.3.6 Constraints on Univariate Predictive Mean Neighborhoods

A univariate imputation was implemented on each variable within each age group after predicted values from the models had been determined. If the respondent was missing the

---

[78]This procedure was implemented using the GENMOD procedure in SAS

value for a particular roster-derived variable, a neighborhood was determined by choosing respondents with predicted values "close to" (within 10% of) the recipient's predicted value.[79] If fewer than 30 donors had predicted values within 10% of the recipient's predicted value, any neighborhood of size greater than zero was allowed. If there were still insufficient donors, the predicted means of the donors were allowed to be within 20% of the recipient's predicted value. The assignment of imputed values for KID17 was restricted within a lower and upper bound based the information that was available in the roster. For example, if a household roster had four members, with two aged 18 or older, one with an age missing, and one with an age under 18, KID17 would be missing. Logically, however, at least one child under age 18 would be in the household, and two adults would be in the household. Hence, the assignment of KID17 in this example would be restricted between the values of 1 and 2. HH65 was restricted within bounds in the same manner.

Likeness constraints were also applied to the imputation of missing values in KID17, HH65, and FAMSKIP. 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 enlarging delta. Similarly, the IRKID17 likeness constraint in the FAMSKIP imputation was loosened after enlarging delta. The likeness constraints and the number of recipients with sufficient donors corresponding to each likeness constraint are summarized in **Appendix I**.

[79]Subsequent theoretical considerations supported the notion that the predicted mean of the donor and recipient had to be as close as possible (within 5%), whereby if no donors were available within 5% of the recipient's predicted mean, the closest donor was chosen and a neighborhood was not used. The 1999 drug use imputations were redone using this criterion.

References

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

Bowman, K., Penne, M., Chromy, J., & Odom, D. (2001, April). *The 1999 National Household Survey on Drug Abuse: Sample design report* (prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 10; RTI/07190.130). Research Triangle Park, NC: Research Triangle Institute.

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

Cox, D.R., & Snell, E.J. (1989). *The analysis of binary data* (2nd ed., Monographs on Statistics and Applied Probability Series). Boca Raton, FL: CRC Press.

Deville, J.C., & Särndal, C.E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*, 376–382.

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

Folsom, R.E., & Singh, A.C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the American Statistical Association, Survey Research Methods Section* (pp. 598-603). Alexandria, VA: American Statistical Association.

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

Kroutil, L.A. (2001a, July). *1999 National Household Survey on Drug Abuse: General principles and procedures for editing drug use data in the 1999 NHSDA computer-assisted interview (for inclusion in the 1999 methodological resource book)* (report prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.195). Research Triangle Park, NC: Research Triangle Institute.

Kroutil, L.A. (2001b, July). *1999 National Household Survey on Drug Abuse: Procedures for editing supplementary self-administered data in the 1999 NHSDA computer-assisted interview (for inclusion in the 1999 methodological resource book)* (report prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.195). Research Triangle Park, NC: Research Triangle Institute.

Kroutil, L.A. (2001c, July). *1999 National Household Survey on Drug Abuse: Procedures for editing interviewer-administered data in the 1999 NHSDA computer-assisted interview (for inclusion in the 1999 methodological resource book)* (report prepared for Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008, Deliverable No. 28; RTI/07190.195). Research Triangle Park, NC: Research Triangle Institute.

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

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

Penne, M.A., Lessler, J.T., Bieler, G., & Caspar, R. (1998). Effects of experimental audio computer-assisted self-interviewing (ACASI) procedures on reported drug use in the NHSDA: Results from the 1997 CAI field experiment. In *1998 Social Statistics Section: Proceedings of the American Statistical Association* (pp. 744-749). Washington, DC: American Statistical Association.

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

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

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

Appendix A

Unweighted Hot-Deck Method of Imputation

This page intentionally left blank.

Appendix A

Unweighted Hot-Deck Method of Imputation

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

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

A.1 Unweighted Sequential Hot Deck

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

A.1.1 Forming Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation procedures were implemented independently within classes defined by the cross of the auxiliary variables. In the main body of the report, these classes were defined by logical and likeness constraints, where classes defined by the likeness constraints could be collapsed if insufficient donors were available, and those defined by logical constraints could not be collapsed, due to the possibility of an inconsistency with preexisting nonmissing values that would result.

A.1.2 Sorting the File

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

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

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

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

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

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

A.1.3 Replacing Missing Values

The file was sorted and then read sequentially. Each time an item respondent was encountered (i.e., the base variable was nonmissing), the base variable response was stored, updating the donor response, and any subsequent nonrespondent encountered received the stored donor response creating the statistically imputed response. A starting value was needed if an item nonrespondent was the first record on a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value.

Note that because the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

A.1.4 Potential Problem

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed there was the risk of several nonrespondents appearing next to one another on the sorted file. To detect this problem in the NHSDA, the imputation donor was identified for every item being imputed. Then, by examining frequencies by imputation donor, one could see whether several nonrespondents were lining up next to one another in the sort. When this problem occurred, sort variables could be added, eliminated, or the order of the variables could be rearranged.

A.2 Unweighted Random Nearest Neighbor Hot Deck

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

A.2.1 Forming Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation procedures were implemented independently within classes defined by the cross of the auxiliary variables. In the main body of the report, these classes were defined by logical and likeness constraints, where classes defined by the likeness constraints could be collapsed if insufficient donors were available, and those defined by logical constraints could not be collapsed, due to the possibility of an inconsistency with preexisting nonmissing values that would result.

A.2.2 Creating a Neighborhood of Potential Donors

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

A.2.3 Randomly Selecting a Donor for the Recipient from the Neighborhood of Donors

From the neighborhood of donors created in the previous step, a single donor was randomly selected whose base variable values would replace those of the recipient. The selection could be conducted as a simple random sample or could incorporate the weights of the potential donors.

Appendix B

Model-Based Method of Item Imputation (PAPI)

This page intentionally left blank.

Appendix B

Model-Based Method of Item Imputation (PAPI)

B.1 Introduction

This appendix describes a model-based approach developed for imputing missing values to categorical survey items. Using this approach, two kinds of probabilities are modeled:

- the probability of a person responding to a question, called the "item response propensity"; and
- the probability that a person answers a question with a particular categorical response, called the "answer category probability."

It is assumed that if any particular question has $J + 1$ possible answer categories, associated with every sample person i will be one item response propensity ρ_i and a set of answer category probabilities $\varphi_{i,j}\ \{j=1,...,J+1\}$ such that

$$\rho_i \geq 0, \quad \forall\ i$$
$$\varphi_{ij} \geq 0, \quad \forall\ i,\ j$$

$$\sum_{j=1}^{J+1} \varphi_{ij} = 1, \ \forall\ i$$

The model-based item imputation procedure has three steps, each of which is described in the following sections.

B.1.1 Step 1: Simultaneously Model Item Response Propensity and Answer Category Probabilities

The first step in the imputation procedure is to model the item response propensity and use these estimated probabilities to reweight the respondents' question-answer data for fitting the answer category probability model. The need to reweight the respondents' question-answer data is motivated by the following set of assumptions/opinions.

Assumption/opinion 1: Survey data are more amenable to accurate modeling of the propensity to collect a data item (item response propensity) than they are to modeling the question's missing values (answer category probabilities). Supporting arguments include the following: (a) The item response indicator is fully observed in the sample, whereas the sample distribution of the answers to the question under consideration is incompletely observed; and (b) fewer covariates with less complex relationships will be required to adequately predict the item response propensity than to model the answer category probabilities.

Assumption/opinion 2: Given the presumed superiority of the item response propensity models ρ_i over the answer category probabilities φ_{ij}, modeling the response propensities and using them to reweight the respondent distribution of answers is likely to accomplish more missing data bias reduction than an imputation method that does not take into account differential item response propensities, particularly when the level of item nonresponse is relatively substantial and is likely to be nonignorable (i.e., the probability of responding depends on the response to the question; see Little & Rubin, 1987). In particular, the item response propensities are used to reweight the respondents' answer category data when fitting the answer category probability model under two possible item response mechanisms, one that is ignorable and one that is nonignorable.

B.1.1.1 Nonignorable Item Response Mechanism. If it is assumed that the conditional response mechanism is nonignorable (conditioned on the covariates used to model the item response propensity), then each item respondent's answer category data are to be reweighted as

$$\left\{\frac{(1-\rho_i)}{\rho_i}\right\}$$

times the respondent's sample weight. This adjustment forces the respondent's reweighted distribution to match the nonrespondent distribution across the covariates used in the response propensity model (see **Section B.3**). Because this adjustment forces the item respondents to be distributed "like" the item nonrespondents (across the item response propensity covariates), when differences between the distributions exist, such differences are compensated for in the respondent reweights used in modeling the answer category probabilities. This reweighting of the data to compensate for the covariate distributional differences between the respondents and nonrespondents should at least partially account for the nonignorable response mechanism.

Note that given the accuracy of the ρ_i predictions, reweighting the observed respondent data by

$$\left\{\frac{(1-\rho_i)}{\rho_i}\right\}$$

yields, on average, the same parameters for the φ_{ij} answer category probabilities that would be achieved if the missing answer category responses could be used exclusively to fit the answer category models. This equivalence "on average" to *nonrespondent*-based answer category probabilities follows from the equivalence in expectation of the weighted

$$\left\{\frac{(1-\rho_i)}{\rho_i}\right\}$$

respondent indicator variable r_i and the nonrespondent indicator $(1 - r_i)$.

B.1.1.2 Ignorable Response Mechanism. In contrast to the conditional nonignorable response mechanism, if it is assumed that the conditional response mechanism is ignorable (conditioned on the covariates used in the item response propensity model), the respondent data are to be reweighted as

$$\frac{1}{\rho_i}$$

times the respondent's sample weight. This adjustment is analogous to treating the response propensity as a final stage, self-selection probability in a Horvitz-Thompson type of estimator. It forces the reweighted respondent distribution to match the total sample distribution for all the item response propensity model covariates.

In this case, it is assumed that the φ_{ij} model parameters are the same on average for the item respondents and nonrespondents. Reweighting the respondent data by

$$\frac{1}{\rho_i}$$

yields model parameter estimates that on average equal those that achieved if both the observed and the missing answer category responses are used to fit the φ_{ij} models.

Note that the complete data-directed weights

$$\frac{1}{\rho_i}$$

(which are used when the conditional response mechanism is assumed ignorable) are generally less variable than their missing data-directed counterparts

$$\frac{(1-\rho_i)}{\rho_i}$$

(which are used when the conditional response mechanism is assumed nonignorable) and therefore lead to less imputation variance. Thus, when deciding on the assumed response mechanism, the nonignorable or nonrespondent-directed solution leads to less biased estimates but an increase in the imputation variance whereas the ignorable or full sample directed solution can lead to more bias in the estimates but a decrease in the imputation variance.

In addition to modeling item response propensity, at this step the answer category probabilities are simultaneously modeled using a polytomous logistic model. The parameters of the polytomous model are estimated using the reweighted respondent data, and the dependent variables in this model are the binary indicator variables for the associated answer categories.

B.1.2 Step 2: Estimate Answer Category Probabilities Among Nonrespondents

The estimated answer category probability model resulting from Step 1 is used to specify imputed categorical response probabilities for the item nonrespondents.

B.1.3 Step 3: Use Answer Category Probabilities to Calculate a Single Categorical Response for the Nonrespondents

After the answer category probabilities are calculated, they are used to randomly assign a single imputed categorical response to each item nonrespondent. This is done using a systematic, probability proportionate to size (PPS) selection technique. This size measure is the answer category probabilities, and this selection technique selects one of the $J + 1$ categorical responses for each item nonrespondent.

This appendix describes each of these three steps in greater detail. **Section B.2** formally presents the models and defines the notation used throughout this appendix. **Section B.3** briefly discusses the method used to estimate the model parameters. **Section B.4** discusses some particular types of covariates used in the models and the reason for using them in the models. And **Section B.5** discusses the methodology used to select a single categorical response for the item nonrespondents using the answer category probabilities.

B.2 Model Specification

The first step in using this model-based imputation procedure is to construct two models: The item response propensity is modeled first, then these propensities are used to reweight the item respondent data in the second model (i.e., the answer category probability model). This section formally presents these models as well as defines the notation used throughout this appendix.

B.2.1 Notation

The indices are defined by the following:

i = sample person, and

j = a single categorical response to the question under consideration, assuming that the question has $J + 1$ categories so $j=1, \ldots, J + 1$.

At the sample unit-level, the variables are defined as follows:

y_i = person i's categorical response to the question under consideration (for simplicity, the categorical responses are assumed to be well-ordered, so $y_i = 1, \ldots, J + 1$);

y_{ij} = indicator variables that equal 1 if $y_i = j$ and equal 0 if y_i is not equal to j;

r_i = item response indicator variable that equals 1 if person i responded to the question under consideration and equals 0 if person i did not respond to the question;

w_i = final analysis weight for person i;

X_i = $(1 \times n_x)$ vector of explanatory variables used in the item response propensity model;

Z_{ij} = $(1 \times n_{zj})$ vector of explanatory variables used in the answer category probability model for response category j $\{j = 1, \ldots, J\}$;

D_i = $(1 \times n_d)$ vector of explanatory variables (typically one-zero subpopulation domain indicators) used in both the item response propensity model and for the answer category probability model (**Section B.4** shows that the components of the vector D_i need to be included in the covariate vectors X_i , Z_{i1} , ..., Z_{ij});

ρ_i = Prob$(r_i = 1 \mid X_i,D_i) = Prob(y_{ij} = 1 \mid Z_{ij},D_i)$ (i.e., item response propensity);

φ_{ij} = Prob$(y_i=j \mid Z_{ij},D_i)$; i.e., answer category probability for category j;

α_i = $(1-\rho_i)/\rho_i$, which are person-level adjustments to the sample weights (w_i) used in the answer category probability model when the response mechanism is conditionally nonignorable for the question under consideration;

ξ_i = $(1/\alpha_i) = \rho_i/(1-\rho_i)$;

λ_i = $1/\rho_i = (1+\alpha_i)$, which are person-level adjustments to the sample weights (w_i) used in the answer category probability model when the response mechanism is conditionally ignorable for the question under consideration;

$$\tau_{ij} = \ln\left\{\frac{\varphi_{ij}}{1-\sum_{j=1}^{J}\varphi_{ij}}\right\}$$

$$= Z_{ij}\,\gamma_{zj} + (v_i D_i)\gamma_{dj} \quad \text{(where } \gamma_{zj},\ v_i,\ \gamma_{dj} \text{ are defined below)}$$

$\tau_i = (1xJ)$ vector of the $\tau'_{ij}s$, $j=1$ ..., J, for each person i

$v_i = X_i\beta_x + [\tau_i{\cdot}D_i]\beta_d$ (where τ_i , D_i and β_d are defined below)

$$= \frac{X_i\ \beta_x + \sum_{j=1}^{J}(Z_{ij}\gamma_{zj})\ (D_i\ \beta_{dj})}{1-\sum_{j=1}^{J}(D_i\ \gamma_{dj})\ \ (D_i\ \beta_{dj})}; \text{ and}$$

$[\tau_i{\cdot}D_i]$ = Kronecker (or direct) product of the vectors τ_i and D_i. $[\tau_i \cdot D_i]$ is a $1 \times (J \times n_d)$ vector that equals:

$$[\tau_i \cdot D_i] = \left\{\tau_{i1}d_{i1},\ldots,\tau_{iJ}d_{i1};\ \tau_{i1}d_{i2},\ldots,\tau_{iJ}d_{i2};\ \ldots\ \ \tau_{i1}d_{in_d},\ldots,\tau_{iJ}d_{in_d}\right\}$$

where

d_{ik} $\{k=1, \ldots, n_d\}$ is the kth element of the vector D_i.

At the sample-level, the quantities are defined as follows:

S_r	=	set of item respondents in the sample;
S_n	=	set of item nonrespondents in the sample;
S	=	$S_r + S_n$ = full sample of unit or questionnaire respondents;
β_x	=	$(n_x \text{ x } 1)$ vector of model parameters associated with the vector X_i;
β_d	=	$((J)(n_d) \text{ x } 1)$ vector of model parameters associated with the vector $\tau_i \cdot D_i$ in the item response propensity model;
β_{dj}	=	$(n_d \text{ x } 1)$ vector of model parameters in β_d associated with the vector $\tau_{ij} \cdot D_i$ in the item response propensity model;
γ_{zj}	=	$(n_{zj} \text{ x } 1)$ vector of model parameters associated with the vector Z_{ij}; and
γ_{dj}	=	$(n_d \text{ x } 1)$ vector of model parameters associated with the vector $v_i\, D_i$.

The constants are defined as follows:

L	=	experimentally fixed constant which will provide a lower bound on $\alpha_i=(1-\rho_i)/\rho_i$ (see the discussion in Section B.2.2);
U	=	experimentally fixed constant to provide an upper bound on $\alpha_i=(1-\rho_i)/\rho_i$ (see discussion in **Section B.2.2**);
A	=	$(U-L) \div [(1-L)(U-1)]$; and
B	=	$L \div (1-L)$.

B.2.2 Item Response Propensity Model

The following item response propensity model is to be estimated:

$$\left\{\frac{1}{A}\right\} \log \left\{\frac{(A-B)\ \xi_i\ -\ (U-1)^{-1}}{1\ +\ B(1-\xi_i)}\right\} = v_i = X_i\beta_x + [\tau_i \cdot D_i]\beta_d \qquad (1)$$

In this model, the independent variables are the vectors X_i and $[\tau_i \cdot D_i]$, the expected value of the dependent variable r_i is ρ_i (recall $\xi_i = \rho_i/(1-\rho_i)$), A and B are fixed constants (functions of the constants U and L), and the parameters of the model to be estimated are β_x and β_d.

Recalling that $\xi_i = \rho_i/(1-\rho_i)$, note that (1) reduces to the unconstrained logistic equation when $L = 0$ and $U = +\infty$ because in this case, the terms $A \Rightarrow 1$, $B \Rightarrow 0$ and $(U-1)^{-1} \Rightarrow 0$. Also, note that when $L = 0$ and $U = +\infty$, the resulting unconstrained logistic equation (1) leads to the logistic weight adjustment factors discussed in Folsom (1991). The logistic weight adjustment procedure discussed in Folsom (1991) was used to derive the nonresponse adjustments for the 1991 to 1994 NHSDA.

Equation (1) can be rearranged to express the nonignorable weight adjustment multiplier $\alpha_i = \xi_i^{-1}$ as

$$\alpha_i = \frac{1-\rho_i}{\rho_i} = \frac{L(U-1) + U(1-L)\exp\left\{-A\left(X_i\beta_x + [\tau_i \cdot D_i]\beta_d\right)\right\}}{(U-1) + (1-L)\exp\left\{-A\left(X_i\beta_x + [\tau_i \cdot D_i]\beta_d\right)\right\}}$$

The constrained logistic model in (1) is an adaptation of a similarly constrained exponential model first suggested by Deville and Särndal (1992) as a method of creating constrained weight adjustment factors that would force weighted sample totals to some other totals (i.e., a generalized raking procedure). From the model (1), notice that for any *i*,

$$\text{As } \left(X_i\beta_x + [\tau_i \cdot D_i]\beta_d\right) \Rightarrow -\infty \text{ then } \alpha_i = (1-\rho_i)/\rho_i \Rightarrow U$$

$$\text{And as } \left(X_i\beta_x + [\tau_i \cdot D_i]\beta_d\right) \Rightarrow +\infty \text{ then } \alpha_i = (1-\rho_i)/\rho_i \Rightarrow L$$

Thus, $L \le \alpha_i \le U$ for any value of $(X_i\beta_x + [\tau_i \cdot D_i]\beta_d)$. Also, note that because $\rho_i = 1/(1+\alpha_i)$, then $(1+U)^{-1} \le \rho_i \le (1+L)^{-1}$.

Recall that L and U are constants in the model and consequently are chosen to bound the resulting adjustments α_i. The purpose of bounding these adjustments is to minimize the imputation variance by minimizing the increase in the original sample weight variability due to the α_i adjustments.

Notice from equation (1) that the independent variable vector $[\tau_i \cdot D_i]$ is a vector defined as a direct product of a vector of explanatory variables D_i and the vectors τ_{ij} (which are described in the next section). The significance of this term in the model is discussed in **Sections B.4** and **B.5**. At this point, note that even though the components of D_i are a subset of the components of X_i, there will not necessarily be a singularity in (1) because $[\tau_i \cdot D_i]$ is a function of D_i and τ_{ij}.

B.2.3 Answer Category Probability Model

The answer category probability models to be estimated are the following:

$$\log \left\{ \frac{\varphi_{ij}}{1 - \sum_{j=1}^{J} \varphi_{ij}} \right\} = \tau_{ij} = Z_{ij}\gamma_{zj} + (v_i D_i)\gamma_{dj} \quad \text{for } j=1, \ldots, J \tag{2}$$

After the φ_{ij} $\{j=1, \ldots, J\}$ are estimated, φ_{iJ+1} is calculated as

$$\varphi_{iJ+1} = 1 - \sum_{j=1}^{J} \varphi_{ij}$$

In this model, the independent variables are the vectors Z_{ij} and $v_i\, D_i$, the expected value of the dependent variables y_{ij} are φ_{ij} $\{j = 1, \ldots, J\}$, and the parameters of the model to be estimated are γ_{zj} and γ_{dj}. Notice that this model is the simple polytomous logistic regression model that yields the desired probabilities of giving answer category j as

$$\text{Prob}\ (y_i = j) = \varphi_{ij} = \frac{\exp\left(Z_{ij}\gamma_{zj} + v_i D_i \gamma_{dj} \right)}{1 + \sum_{j'=1}^{J} \exp\left(Z_{ij'}\gamma_{zj'} + v_i D_i \gamma_{dj'} \right)} \quad \text{for } j=1, \ldots, J$$

And

$$\text{Prob}\ (y_i = J+1) = \varphi_{i\,J+1} = \frac{1}{1 + \sum_{j'=1}^{J} \exp\left(Z_{ij'}\gamma_{zj'} + v_i D_i \gamma_{dj'} \right)}$$

As with the item response propensity model, the answer category probability model includes independent variables *(v_i D_i)* that are a function of the estimated item response propensities (ρ_i), via α_i, which are estimated in equation (1). The significance of the $v_i D_i$ term is discussed in **Sections B.4** and **B.5**. As with the $[\tau_i \cdot D_i]$ term in (1), it is significant to note that $v_i D_i$ is a function of D_i and ρ_i, so even though the components of D_i are also in Z_{ij} there will not necessarily be a singularity in equation (2).

B.3 Estimating the Model Parameters

This section briefly summarizes the method used to estimate the parameters in the item response propensity model (equation 1) and the answer category probability model (equation 2).

B.3.1 Raking Equations from the Item Response Propensity Model

As mentioned in **Section B.2.2**, equation (1) was adapted from a formula first suggested by Deville and Särndal (1992) as a method of producing constrained weight adjustment factors that would adjust the sample weights to some other specified totals. They arrived at equation (1) by minimizing a particular distance function subject to a set of constraints. The function they minimized is a measure of the distance between the adjusted weight and the initial sample weight. They minimized this distance in order to minimize the bias and any increase in the variance that results from adjusting sample weights. For the nonresponse application, their constraints required that the $w_i(1+\alpha_i)$ respondent-weighted X_i element totals must equal the associated full sample totals:

$$\sum_{i \in S_r} w_i(1+\alpha_i)\ X_i \ -\ \sum_{i \in S} w_i\ X_i \ =\ \tilde{0}$$

To achieve the equivalence between the alternative domain prevalence estimators presented in **Section B.4**, a set of analogous constraint equations was added for a vector of constructed variables derived from the answer category probabilities φ_{ij}. Specifically, the following constraint equations were added:

$$\sum_{i \in S_r} w_i(1+\alpha_i)\ [\varphi_i \bullet D_i] \ -\ \sum_{i \in S} w_i\ [\varphi_i \bullet D_i] \ =\ \tilde{0}$$

Using the notation specified in this appendix, solving their constrained, minimization problem reduces to solving for β_x and β_d in the following set of calibration (or raking) equations:

$$\sum_{i\varepsilon S_r} w_i\ \alpha_i\ X_i\ -\ \sum_{i\varepsilon S_n} w_i X_i\ =\ \tilde{0}$$

$$\sum_{i\varepsilon S_r} w_i\ \alpha_i\ [\varphi_i \cdot D_i]\ -\ \sum_{i\varepsilon S_n} w_i [\varphi_i \cdot D_i]\ =\ \tilde{0}$$

Or equivalently,

$$\left.\begin{aligned} &\sum_{i\varepsilon S} w_i\ [r_i\alpha_i - (1-r_i)]\ X_i\ =\ \tilde{0} \\ &\sum_{i\varepsilon S} w_i\ [r_i\alpha_i - (1-r_i)]\ [\varphi_i \cdot D_i]\ =\ \tilde{0} \end{aligned}\right\} \qquad (3)$$

Note that X_i and $[\varphi_i \cdot D_i]$ are vectors with n_x and (J x n_d) components, respectively; β_x and β_d are vectors with n_x and (J x n_d) components, respectively; and n_x + (J x n_d) equations are presented in equation (3). Thus, the n_x + (J x n_d) unknowns $\{\beta_x, \beta_d\}$ are to be solved in the equations of (3).

Intuitively, the calibration equations (3) indicate that β_x and β_d are to be solved in order to derive sample weight adjustment factor (i.e., the α_i's) that will adjust the sample weights of the item respondents so that they reproduce the nonrespondent-weighted sum of the X_i and $[\varphi_i \cdot D_i]$ covariate vectors. To illustrate this, suppose that one of the elements of the vector X_i is a one-zero indicator for 12 to 17 year olds in the sample (the indicator = 1 if person *i* is 12 to 17 years old, 0 otherwise). Then, the first set of calibration equations in (3) implies that the weight adjustment α_i to be solved would make the reweighted respondent count of 12 to 17 year olds equal to the originally weighted count of nonrespondents 12 to 17 years old.

Note that in these raking equations, one would expect to see $[\tau_i \cdot D_i]$ in the second set of equations of (3) because this is a factor in the definition of α_i; however, this factor has been replaced with $[\varphi_i \cdot D_i]$. This substitution was done because an equivalence needed to be obtained between four important estimators (see **Section B.4**). Furthermore, obtaining strict equality between the reweighted respondent [$\tau_i \cdot D_i$] totals and the corresponding full sample totals was not necessary. The components of D_i are a subset of X_i, and equality between the reweighted respondent X_i totals with the corresponding full sample totals is already being achieved with the first set of equations in (3).

B.3.2 Score Functions from the Answer Category Probability Model

The parameters of the answer category probability models (2) are estimated using an adaptation of the method of design weighted or pseudo-maximum likelihood. Using this method, the value of γ_{zi} and γ_{dj} to be found maximizes the pseudo-likelihood function:

$$L(\gamma) = \prod_{i \varepsilon s_r} \left\{ \prod_{j=1}^{J} \left[\varphi_{ij} \right]^{\alpha_i w_i y_{ij}} \left[1 - \sum_{j=1}^{J} \varphi_{ij} \right]^{\alpha_i \, w_i \, (1 - \sum_{j=1}^{J} y_{ij})} \right\} \qquad (4)$$

where

$$\varphi_{ij} = \frac{\exp\left[Z_{ij} \ \gamma_{zj} + v_i D_i \gamma_{dj} \right]}{1 + \sum_{j'=1}^{J} \exp\left[Z_{ij'} \ \gamma_{zj'} + v_i D_i \gamma_{dj'} \right]} \qquad \text{for } j = 1, \ ..., \ J$$

$$\varphi_{iJ+1} = \frac{1}{1 + \sum_{j'=1}^{J} \exp\left[Z_{ij'} \gamma_{zj'} + v_i D_i \gamma_{dj'} \right]}$$

The problem of maximizing this pseudo-likelihood function reduces to finding the roots of the partial derivatives of (4) with respect to the γ_{zj} and γ_{dj}; namely,

If an ignorable item response mechanism is assumed, then α_i should be replaced with $\lambda_1 = (1 + \alpha_1) = p_1^{-1}$ in equation (5).

The equations (5) are called the score functions or likelihood equations. Note for j = 1, ..., J: Z_{ij} and $\nu_i D_i$ are vectors with n_{zj} and n_d components, respectively; γ_{zj} and γ_{dj} are vectors with

$$\left.\begin{array}{l}\sum_{i\varepsilon S_r} \alpha_i w_i \varphi_{ij} Z_{ij} - \sum_{i\varepsilon S_r} \alpha_i w_i y_{ij} Z_{ij} = \tilde{0} \ , \quad \text{for } j = 1, \ ..., \ J \\ \\ \sum_{i\varepsilon S_r} w_i \varphi_{ij} D_1 - \sum_{i\varepsilon S_r} w_i y_{ij} D_i \ = \ \tilde{0} \ , \quad \text{for } j = 1, \ ..., \ J\end{array}\right\} \qquad (5)$$

n_{zj} and n_d components; and $n_{zj}+n_d$ equations are presented in (5). Thus, for each *j*, the $n_{zj}+n_d$ unknowns $\{\gamma_{zj}, \ \gamma_{dj}\}$ are to be solved in the $n_{zj}+n_d$ equations of (5).

Note the similarities between the score functions (5) and the calibration equations (3). In (5), γ_{zj} and γ_{dj} are to be solved in a function that will force an equality between the actual observed response y_{ij} and the estimated probabilities φ_{ij} among the item respondents. Essentially, the only difference between (3) and (5) are in the functions of the variables to be estimated: α_i, a function of the β's in equation (3), and φ_{ij}, a function of the γ's in equation (5).

As with the raking equations (3), note that one would expect to see the additional multiplier $\alpha_i \nu_i$ in the second set of equations displayed in (5). This would follow from the $\nu_i D_i$ term in the definition of φ_{ij}. The ν_i is treated as if it was $(1 \div \alpha_i)$ instead of $\ln(1 \div \alpha_i)$ in the unconstrained case. The absence of this term permits algebraic equality between four estimators, which is further discussed in **Section B.4**.

B.3.3 Parameter Estimation

To estimate the parameter vectors β_x, β_d in the raking equations (3) and the parameter vectors γ_{zj}, γ_{dj}, j = 1, ..., J in the score functions (5), the Newton-Raphson algorithm was used to obtain a simultaneous solution. These parameters needed to be solved for simultaneously because of the terms that these equations have in common. Note that the term τ_i in (3) is a function of the answer category probabilities (2). Further, α_i is not only an explicit factor in (5), but also note that ν_i is a function of α_i through equation (1).

If $\Psi \ = \ \{\beta_x; \ \beta_d; \ \gamma_{zj} \ j=1,...,J; \ \gamma_{dj} \ j=1,...,J\}$, $F(\Psi)$ denotes the system of equations presented by (3) and (5) and $J^{-1}(F(\Psi))$ equal the inverse Jacobian of F (i.e., the inverse of the matrix of partial derivative of F with respect to Ψ), the vector Ψ is found by iteration

$\Psi^o \quad = \quad 0$ at iteration 0, and

$\Psi^{K+1} \quad = \quad \Psi^k - J^{-1}[F(\Psi^k)]\ F(\Psi)^k)$ at iteration K=1,2,3, ...

Generally, Ψ^k is iterated until $|F(\Psi)| < \varepsilon$ (usually set at .00001).

B.4 Looking at the D_i Term in the Models

The significance of the D_i vector in the item response propensity and answer category probability models is that for any particular categorical response *j* to the question under consideration, and for each component d_i of the vector D_i, the following four estimators will be equal for the subpopulation fraction that would respond with answer category *j*:

Estimator 1: The Mean of Respondent y_{ij} Values, Weighted with λ_i Adjusted Weights

$$\frac{\sum_{i \varepsilon S} w_i \lambda_i r_i d_i y_{ij}}{\sum_{i \varepsilon S} w_i \lambda_i r_i d_i}$$

Estimator 2: The Mean of Respondent y_{ij} Values and Predicted Nonrespondent φ_{ij} Probabilities, Weighted with The Original Sample Weights

$$\frac{\sum_{i \varepsilon S} w_i d_i \ \{ r_i y_{ij} + (1 - r_i) \varphi_{ij} \}}{\sum_{i \varepsilon S} w_i d_i}$$

Estimator 3: The Mean of Predicted φ_{ij}, Weighted with the Original Sample Weights

$$\frac{\sum_{i \varepsilon S} w_i d_i \varphi_{ij}}{\sum_{i \varepsilon S} w_i d_i}$$

Estimator 4: The Mean of Predicted Respondent φ_{ij} Probabilities, Weighted with λ_i Adjusted Weights

$$\frac{\sum_{i \epsilon S} w_i \lambda_i r_i d_i \varphi_{ij}}{\sum_{i \epsilon S} w_i \lambda_i r_i d_i}$$

To motivate the benefit of achieving equality between these four estimators, note the following:

- Assuming that the item response propensity model is correct (which implies ignorable nonresponse), the λ_i reweighted mean of the nonmissing answer category vectors y_i (Estimator 1) is free of item nonresponse bias. Note that the correctness of the model for the item response propensity implies that the y_i vectors are missing-at-random given X_i and $[\tau_i \cdot D_i]$, (Little & Rubin, 1987) because ρ_i; therefore, the distribution of r_i is independent of the answer category vector y_i conditional on X_i and $[\tau_i \cdot D_i]$.

- Estimator 2, the estimator that uses the imputed φ_i for the item nonrespondents, is the basis for the randomized imputation procedure (discussed in **Section B.5**). Because Estimator 2 is algebraically equivalent to Estimator 1 (shown later in this section), the randomized imputation estimator also is free from item nonresponse bias assuming ignorable nonresponse. Note that this unbiasedness derives from the correctness of the item response propensity model ρ_i and the algebraic equivalence of Estimators 1 and 2. The φ_i=E{y_i|X_i, ν_i D_i} model is not required to be correct for the answer category probabilities.

- Estimator 3 is a double-sampling regression type of estimator with the questionnaire respondents constituting the first phase sample. The item respondents are viewed as the second phase sample, with the ρ_i treated as known self-selection probabilities for the second phase. This estimator is a generalization of the familiar linear regression estimator for double-sampling designs where φ_{ij} is the analog of the linear least squares predictor for y_{ij} derived from the covariates observed in Phase I. It can be shown that the equivalence of Estimators 3 and 2 yields a useful variance form for Estimator 2 in terms of multiple $R^2_{\varphi|x}$ measure.

- The inclusion of the $[\tau_i \cdot D_i]$ term in the item response propensity model gives the equality between Estimators 3 and 4 given the calibrations equations (3).

- Using the weight adjustments α_i (or λ_i) in the answer category models, along with the $\nu_i D_i$ term, gives the equality between Estimators 1 and 4 and Estimators 2 and 3. This equality establishes the desired equivalence between Estimators 1 and 2 given the identity between 3 and 4 claimed in the previous statement.

In this section, it is shown that these four estimators are algebraically equal, assuming the item response mechanism is nonignorable (i.e., using the nonrespondent directed solution). A similar argument can be presented that would show why these four estimates are equal assuming the item response mechanism is ignorable (i.e., using the respondent-directed solution).

B.4.1 Estimator 1 = Estimator 4

The following shows that the equality between these estimators is achieved given that $d_i \in D_i$ is a component in Z_{ij} for all j. The result follows from the identities stipulated by the score functions in equation (5).

Suppose d_i is a component of D_i. Then

$$\sum_{i \varepsilon S} w_i \lambda_i r_i d_i y_{ij}$$

$$= \sum_{i \varepsilon S} w_i (\alpha_i + 1) r_i d_i y_{ij}$$

$$= \sum_{i \varepsilon S} w_i \alpha_i r_i d_i y_{ij} + \sum_{i \varepsilon S} w_i r_i d_i y_{ij}$$

$$= \sum_{i \varepsilon S} w_i \alpha_i r_i d_i \varphi_{ij} + \sum_{i \varepsilon S} w_i r_i d_i \varphi_{ij}$$ from likelihood equations (5) and because $d_i \varepsilon D_i$ and $d_i \varepsilon Z_{ij}, \forall_j$

$$= \sum_{i \varepsilon S} w_i \lambda_i r_i d_i \varphi_{ij}$$

B.4.2 Estimator 2 = Estimator 3

The following shows that this equality is achieved using an argument similar to what was used to show that Estimator 1 equals Estimator 4 (i.e., equality is achieved because d_i is a component of the vector Z_{ij} for all j, and because of the score functions (5).

$$\sum_{i \varepsilon S} w_i d_i \{ r_i y_{ij} + (1 - r_i) \varphi_{ij} \}$$

$$= \sum_{i \varepsilon S} w_i d_i r_i y_{ij} + \sum_{i \varepsilon S} w_i d_i (1 - r_i) \varphi_{ij}$$

$$= \sum_{i\varepsilon S} w_i d_i r_i \varphi_{ij} \quad + \quad \sum_{i\varepsilon S} w_i d_i (1-r_i)\varphi_{ij} \qquad \text{from the likelihood equations (5)}$$

$$= \sum_{i\varepsilon S} w_i d_i \varphi_{ij}$$

B.4.3 Estimator 3 = Estimator 4

To show this equality, first note that the denominators of the two estimators are equal due to the calibration equations (3) and because d_i is a component of the vector X_i:

$$\sum_{i\varepsilon S} w_i \lambda_i r_i d_i$$

$$= \sum_{i\varepsilon S} w_i (1 \ + \ \alpha_i) r_i d_i$$

$$= \sum_{i\varepsilon S} w_i r_i d_i \quad + \quad \sum_{i\varepsilon S} w_i \alpha_i r_i d_i \qquad \text{from the calibration equations (3)}$$

$$= \sum_{i\varepsilon S} w_i d_i$$

And for the same reasons, the numerators are equal.

$$\sum_{i\varepsilon S} w_i \lambda_i r_i d_i \varphi_i$$

$$= \sum_{i\varepsilon S} w_i (\alpha_i + 1) r_i d_i \varphi_{ij}$$

$$= \sum_{i\varepsilon S} w_i r_i d_i \varphi_i \quad + \quad \sum_{i\varepsilon S} w_i \alpha_i r_i d_i \varphi_{ij}$$

$$= \sum_{i\varepsilon S} w_i r_i d_i \varphi_{ij} \quad + \quad \sum_{i\varepsilon S} w_i (1-r_1) \varphi_{ij} d_i \qquad \text{from calibration equations (3)}$$

$$= \sum_{i\varepsilon S} w_i d_i \varphi_i$$

B.4.4 Estimator 1 = Estimator 2

This crucial identity is the obvious consequence of the previous three. It has been shown that

Estimator 1 = Estimator 4

Estimator 3 = Estimator 4

Estimator 2 = Estimator 3.

So it follows that

Estimator 1 = Estimator 2.

B.5 Using the Answer Category Probabilities to Generate a Single Categorical Response for the Item Nonrespondents

The final step in this model-based imputation procedure is to take the predicted answer category probabilities from the polytomous logistic model and select a single categorical response for each nonrespondent. This step is done mainly for the convenience of data analysts. Note that **Section B.4** presents four population estimators based on the answer category probabilities. Unfortunately, survey tabulation software typically does not allow for easy computation of population estimates based on answer category probabilities. Thus, to simplify the data analysis, the estimated φ_{ij} is used to select a single answer category for each nonrespondent.

To select an answer category for each nonrespondent, a probability proportionate to size (PPS) algorithm is used with the φ_{ij} as the size measure. This procedure first identifies a subsample of the item nonrespondents who have their answer category set equal to the first response category (say, $j = 1$). This subsample is then removed from the initial set of nonrespondents, and the conditional probabilities of responding to category $j = 2, ..., J + 1$ are calculated for each remaining nonrespondent. This process of selecting a subset of nonrespondents for category j and calculating the conditional probability of responding to categories $j + 1, ...,J + 1$ for each remaining nonrespondent continues until all nonrespondents are assigned to a single category.

The following describes this algorithm in greater detail. Step 1 entails simply sorting the set of nonrespondents in some well-defined order. Step 2 involves more detailed operations. For $j = 1, ..., J$, do the following:

- Assign θ_j to some randomly chosen number from a uniform (0,1) distribution.

- Assign the variable $S_j = 0$. The variable S_j will store the accumulated size measure for response category j as the algorithm proceeds through the set of nonrespondents.

- Sequentially read the set of nonrespondents, and for each nonrespondent do the following. For j = 1, ..., J:

$$\text{If } S_j < \theta_j < S_j + \varphi_{ij}, \text{ then}$$

$$\theta_j = \theta_{j+1},$$

$$S_j = S_j + \varphi_{ij},$$

$$\varphi_{ik} = 0 \text{ for } k = j + 1, ..., J.$$

At this branch in the algorithm, nonrespondent i is assigned to category j. Otherwise, $S_j = S_j + \varphi_{ij}$ and

$$\varphi_{ik} = \frac{\varphi_{ik}}{\sum_{k'=j+1}^{J+1} \varphi_{ik'}} \qquad \text{for } k=j+l,...,J+1.$$

At this branch in the algorithm, nonrespondent i was not selected for category j, so the conditional probabilities of being in category j + 1, ..., J + 1 are calculated (conditioned on the nonrespondent not being selected for category 1, ..., j).

At the end of this loop, if the nonrespondent was not selected for category j = 1, ..., J, then it is asigned to category J + 1.

Using the probabilities as the size measure in this algorithm provides the following: where $\tilde{y}_{ij}$ is the assigned y_{ij} from the randomization.

$$\sum_{i \varepsilon S_n} \tilde{y}_{ij} \simeq E\left\{\sum_{i \varepsilon S_n} \tilde{y}_{ij}\right\} = \sum_{i \varepsilon S_n} \varphi_{ij} \quad , \quad \forall\, j$$

The main shortcoming of any randomization algorithm is that although the number of total nonrespondents assigned to category j will roughly equal the sum of the category j probabilities, close correspondence between the sample-weighted sums for survey domains of interest may not be achieved by any particular randomization. In other words, due to randomization variance, the following estimators may not be very close to one another:

$$\sum_{i \varepsilon S_n} w_i \eta_i \tilde{y}_{ij} \simeq \sum_{i \varepsilon S_n} w_i \eta_i \varphi_{ij} \qquad (6)$$

where η_i is some vector whose components are a subset of the components in the vectors X_1, Z_{ij}, and D_i. Typically, η_i will contain the primary domain variables of interest for the question under consideration.

To compensate for this disadvantage, multiple completed datasets are built by selecting multiple categorical responses for each nonrespondent using the above selection technique. Then by examining the equality of (6) for each completed dataset (after statistically testing the equality using a simple chi-squared test), the best randomization will be chosen from among the multiple imputations.

Appendix C

Technical Details About the Generalized Exponential Model (GEM)

This page intentionally left blank.

Appendix C

Technical Details About the Generalized Exponential Model (GEM)

C.1 Distance Function

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

$$\Delta(w,d) = \sum_{k \in s} \frac{d_k}{A_k} \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\} \tag{C1.1}$$

where $a_k = w_k / d_k$, $A_k = (u_k - \ell_k) / (u_k - c_k)(c_k - \ell_k)$, and ℓ_k, c_k, u_k are prescribed real numbers. Let $\mathbf{T}_x$ denote the p-vector of control totals corresponding to predictor variables (x_1, ..., x_p, say). Then the calibration constraints for the above minimization problem are

$$\sum_{k \in s} x_k \, d_k \, a_k = T_x \tag{C1.2}$$

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

$$a_k(\lambda) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp\{A_k \boldsymbol{x_k}' \boldsymbol{\lambda}\}}{(u_k - c_k) + (c_k - \ell_k) \exp\{A_k \boldsymbol{x_k}' \boldsymbol{\lambda}\}} \tag{C1.3}$$

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 (C1.3) that

$$\ell_k < a_k < u_k, \; k = 1, .., n \tag{C1.4}$$

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

$$\Delta(w,d) = \sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k(a_k - 1) \tag{C1.5}$$

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

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

C.2 GEM Adjustments for Extreme Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (C1.3) 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 = $\text{med}\{d_k\} \pm 3 * \text{IQR}$ if d_k is an outlier, where the quartiles for the weights are defined with respect to a suitable design-based stratum. Then with GEM for outlier treatment, $\ell_k = 1, c_k = c = 1 + \Sigma_{s**}(d_k - b_k) / \Sigma_{s*} d_k$ and $u_k = u > c$ can be chosen for nonoutliers, and the outliers are held fixed at their winsorized values, where s_* denotes the subsample of nonoutliers, and s_{**} the subsample of outliers.

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

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

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

C.3 Newton-Raphson Steps

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

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

where

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

Then at the Newton-Rahpson iteration ν, the value of the p-vector $\boldsymbol{\lambda}$ is adjusted as

$$\boldsymbol{\lambda}^{(\nu)} = \boldsymbol{\lambda}^{(\nu-1)} + (\boldsymbol{X}'\ \boldsymbol{\Gamma}_{\varphi,\nu-1}\boldsymbol{X})^{-1}\ (T_{\boldsymbol{x}} - \hat{T}_{\boldsymbol{x}}^{(\nu-1)}) \tag{C3.1}$$

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

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

C.4 Scaled Constrained Exponential Model

In previous NHSDAs, constrained exponential models (CEM) were used for ps and scaled CEM for nr adjustments. The CEM refers to the logit model of Deville and Särndal (1992) in which lower and upper bounds do not vary with *k* (i.e., $\ell_k = \ell$, $u_k = u$, and $c_k = c = 1$ such that $\ell < 1 < u$). Thus, it is a special case of GEM. For the nr adjustment, Folsom and Witt (1994) modified CEM estimating equations by a scaling factor (ρ^{-1}: inverse of the overall response propensity) such that $1 < \rho^{-1} a_k < \rho^{-1}\, u$. This implies that by choosing ℓ in CEM as ρ, one can ensure that the scaled adjustment factor for nonresponse is at least 1.

Appendix D

Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

This page intentionally left blank.

Appendix D

Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

D.1 Introduction

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

D.2 Overview

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

The new method, called Predictive Mean Neighborhoods (PMN), is a combination of two commonly used imputation methods: a non-model-based hot deck (nearest neighbor), and a modification of the model-assisted predictive mean matching method of Rubin (1986). PMN enhances the predictive mean matching (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 nearest neighbor hot deck (NNHD) (Little & Rubin, 1987, p. 65). With this method, donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (the donor has complete data, the recipient does not). A donor set deemed close to the recipient with respect to a number of covariates is used to select a donor at random. For the NHSDA, the set of covariates typically would include demographic variables as well as some other nonmissing drug use variables. To further ensure that a donor matches the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with drug use, such as age categories, can be used to restrict the set of donors. Furthermore, other restrictions involving outcome variables can be imposed on the neighborhood. Note that in NNHD, unlike sequential hot deck, a distance function is used to define closeness between the recipient and a donor. So there is less of a problem of sparseness of the donor class, but the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

The predictive mean matching method (PMM) 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

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

2. 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 predictive mean matching method is not applicable to discrete variables because the distance function between recipient's predictive mean (which takes continuous values) and donor's outcome value (which takes discrete values) is not well defined.

D.2.2 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

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

The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be tedious in practice because of the computational problems due to the sheer number of model parameters, and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, none of the existing solutions take the survey design into account because of the obvious problem of specifying the probability distribution underlying survey data. However, in the application of the multivariate predictive mean neighborhood imputation (MPMN) to the 1999 NHSDA, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case) such that variables modeled earlier on in the hierarchy have a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for use of univariate techniques for fitting multivariate (but conditional) predictive means.

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

D.3 Outline and Description of Method

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

D.3.1 Step 1: Hierarchy Definition

The first step is to determine the order in which variables are modeled, so that variables early in the hierarchy may be used for modeling the conditional predictive mean (i.e., they have the potential to be part of the set of covariates for variables later in the hierarchy). Note that not all variables in the hierarchy may be missing for a particular incomplete record. Nevertheless, models are to be developed for all the variables in a univariate fashion for reasons mentioned earlier. For example, in the drug modules in the CAI sample of the 1999 NHSDA, different drugs needed to be modeled, with different measures of drug use for each drug. It was therefore necessary to determine the order in which the combination of drugs and drug use measures were to be handled. Using the sequence of variables determined by this step, the procedure involved cycling through Steps 2 through 5, and sometimes 6. In the application of the PMN to the NHSDA, the order of imputation for drugs was determined by considering such factors as the level of stigma associated with the drugs, the level of "missingness" in the data (see Appendix I), and the degree to which one set of drugs could be used as predictors for other drugs. The order of drugs was given by cigarettes, smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, crack, and heroin. The order of drug use measures imputed was determined based on the natural hierarchy of the variables: lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past 30 days, and age of first use.

For each variable, Steps 2 through 5 are to be followed.

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

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

D.3.3 Step 3: Sequential Hierarchical Modeling

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

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

Once the model has been fitted, the predictive means for item respondents and item nonrespondents are to be calculated using the model coefficients. For models with a multivariate predictive mean vector (such as with a polytomous logit model), a single element out of that vector must be chosen, so that each respondent has exactly one predictive mean value.[1] This predictive mean is the matching variable in a random nearest neighbor hot deck. It can come directly from the model, it can be adjusted to account for the conditioning on the time period, or (if it is the predicted value based on a model with a transformed response variable) it can be back-transformed to the original units.

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

The pool of donors is to be further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a "logical constraint," is given by age at first crack use, which must not be younger than age at first cocaine use. Other constraints, called "likeness constraints," are placed on the pool of donors to make the attributes of the neighborhood as close to that of the recipient as possible. For example, for age at first use, the age of the donor and the age of the recipient are restricted to be the same whenever possible, and the donor and recipient must come from States with similar usage patterns. A small value of delta could also be thought of as a likeness constraint. Whenever insufficient donors are available to meet the likeness constraints, including the preset small value of delta, the constraints are to be loosened in priority order

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

according to their perceived importance. As a last resort, if an insufficient number of donors are available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor is to be found using a sequential hot deck, where matching is to be done on the predictive mean. (Even though weights would not be used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process because weighting would already have been incorporated in the calculation of the predicted mean.)

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

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

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

The assignment step is multivariate if several response variables are associated with a single predictive mean, provided more than one of those response variables is missing. In that case, all of the missing values is to be imputed using the same donor. If there is more than one response variable associated with a single predictive mean, but not all of them are missing, only the missing values are to be replaced by those of the donor. The resulting imputed values are provisional if a multivariate neighborhood (MPMN) step is called for; otherwise, these values are final.[2]

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

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

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

With MPMN, the neighborhood is defined based on a vector of predictive means rather than from a single predictive mean as in the univariate case. This vector may encompass a subvector of predictive means from a single categorical model (as with a polytomous logit model), in addition to scalar predictive means from any number of models with continuous response variables. For each item nonrespondent, a distance is to be calculated between the elements of this vector of predictive means where the observed values are missing, and the corresponding elements of the vector for every item respondent. To make all elements of the vector conditional on the same usage status in the full predictive mean vector, predictive means that were calculated on the basis of past year and past month users are to be adjusted to account for the probability that a respondent is a past year user or a past month user. For example, in the CAI sample of the 1999 NHSDA, the full predictive mean vector for alcohol included the following elements:

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

2. *recency, past year, not past month*: *P* (past year but not past month alcohol user | lifetime alcohol user);
3. *12-month frequency*: P (the respondent used alcohol on a given day in the past year | past year user of alcohol) * *P* (past year user of alcohol | lifetime alcohol user)[3];

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

---

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

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

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

P (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).[5]

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

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

where μ_R refers to the predictive mean (sub-)vector for a given item respondent, and μ_{NR} is the predictive mean (sub-)vector for a given item nonrespondent. The matrix Σ is the variance-

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

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

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

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

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

D.4 Comparison of PMN with Other Available Methods

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

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

- **Relative importance of covariates is determined by standard estimating equation techniques.** In other words, there are objective criteria based on methodology, such as regression, that quantify the relationship between a given covariate and the response variable, in the presence of other covariates. Thus the response variable itself is indirectly used to determine donors.

- **The problem of sparse neighborhoods is considerably reduced which makes it easier to implement restrictions on the donor set**. Because the distance function is defined as a continuous function of the predictive mean, it is possible to find donors arbitrarily close to the recipient. Thus, it is less likely to have the problem of sparse neighborhoods for hot decking. Moreover, having sufficient donors in the neighborhood allows for imposing extra constraints on the donor set, which would have been difficult to incorporate directly in the model.

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

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

- **The choice of donor can be made random by choosing delta large enough such that the neighborhood is of a size greater than 1.** Under the assumption that the recipient and the candidate donors in the neighborhood have approximately equal means, the random selection allows us to mimic the case where the error distribution has mean zero. 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 to other model-based methods, discrete and continuous variables can be handled jointly and relatively easily in MPMN by using the idea of univariate (conditional) modeling in a hierarchical manner. In MPMN, one can objectively assign differential weights to different elements of the predictive mean vector depending on the variability of predictive means in the data set 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 E

Race and Hispanic-Origin Group Alpha Codes

This page intentionally left blank.

Appendix E

Race and Hispanic-Origin Group Alpha Codes

To reduce the amount of statistical imputation necessary to create the imputation-revised race and Hispanic-origin variables for the computer-assisted interviewing (CAI) sample of the 1999 National Household Survey on Drug Abuse (NHSDA), the race and Hispanic-origin group alpha-specify dictionaries used in prior NHSDAs were expanded, as were the procedures used to assign the large number of other-specify responses to the categories used for these variables. As discussed in **Chapter 4**, many respondents provided a race in the alpha-specify response to the Hispanic-origin group question, and vice-versa, so responses to both questions were examined in the creation of each variable. The following 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.

E.1 Race

In a change from the paper-and-pencil interviewing (PAPI) questionnaire (1999 and past years), respondents were permitted to select more than one race in the 1999 CAI questionnaire. There also was a follow-up question asking respondents who selected multiple races in the first question to select from among those chosen the single race that best described them. As in past years (and the 1999 PAPI), respondents had an opportunity to specify a race not included in the question by responding "other," either as the sole race chosen in the first question, or as the race that best described them if "other" was among multiple races chosen in the first question. The CAI race questions used in 1999 are as follows:

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

1	White
2	Black/African American
3	American Indian 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
13 Other (Specify)

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

(Choose from among responses to QD05)

E.1.1 Race Alpha Responses

The other-specify responses were examined when (a) "other" was selected as a race in QD05,[1] or (b) no race was given in response to QD05, but a race category was given as an other-specify response to the Hispanic-origin group question (QD04). In such cases, if a respondent provided a valid alpha-specify response when asked, that response was used in order to assign a value of EDRACE, the base variable for imputing IRRACE (see **Chapter 4**), as follows:

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

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

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

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

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

5. If a respondent indicated only an Hispanic-origin group in response to the race other-specify question, then 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, and Cubans and Central or South Americans combined.

6. For certain countries of origin given in the other-specify responses, race was randomly assigned using census data for those countries. In many cases, a small percentage of respondents from a given country were left to be statistically imputed. The following is a list of the countries treated this way and the percentages assigned to each race:[2]

   - Dominican Republic: 84% black, 16% white, 0% statistically imputed;
   - Caribbean and West Indies: 80% black, 14% Asian, 6% statistically imputed;
   - Belize: 55% American Indian, 37% black, 8% statistically imputed;
   - Guyana: 51% Asian, 43% black, 6% statistically imputed;
   - Suriname: 52% Asian, 31% black, 17% statistically imputed;
   - Trinidad and Tobago: 57% black, 40% Asian, 3% statistically imputed;
   - Jamaica: 91% black, 9% statistically imputed;

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

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

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

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

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

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

E.1.2 Assigning a Race When Multiple Races Were Selected

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

E.1.3 Race Dictionary Codes

Codes were assigned to respondents based either on their response to the first 12 categories of QD05 (codes 1 to 12), or on their race alpha-specify responses (codes 21 to 985). Codes 21 to 32 are equivalent to codes 1 to 12, except that the race identification was obtained from the alpha-specify responses. The values of EDRACE were obtained using these codes (see **Section E.1.2**), which are presented below:

1	White
2	Black/African American
3	American Indian or Alaska Native
4	Native Hawaiian
5	Other Pacific Islander
6	Chinese
7	Filipino
8	Japanese
9	Asian Indian
10	Korean
11	Vietnamese
12	Other Asian
21	White (includes Arab, Turkish, Armenian, Jewish)
22	Black/African American (includes Haiti, St. Vincent, Dominica)
23	American Indian or Alaska Native (includes mestizo)
24	Native Hawaiian
25	Other Pacific Islander
26	Chinese
27	Filipino
28	Japanese
29	Asian Indian
30	Korean
31	Vietnamese
32	Other Asian (includes Iran, Kurd, Afghan)
33	Asian nonspecific
34	Guamanian
41	Hispanic (nonspecific, race not given)
42	Mexican
43	Puerto Rican
44	Central or South American (excludes Belize/Guyana/Suriname)
45	Cuban
46	Dominican Republic (Santo Domingo)
47	Dominica (Roseau)
48	Dominican (Dominican Republic vs. Dominica not clear)
49	Caribbean/West Indies
50	Belize
51	Guyana
52	Suriname
53	Trinidad and Tobago
54	Jamaica
55	Virgin Islands (St. Thomas, St. Croix), Bahamas
80	United Kingdom
81	Portugal/European Spanish
82	Spanish, maybe European
83	Other Western Europe
84	Middle East/Israel/North Africa
85	Canada
86	New Zealand
87	South Africa
88	Australia
101	Part Hispanic, part white
102	Part Hispanic, part black
103	Part Hispanic, part American Indian
104	Part Hispanic, part Asian
105	Part Hispanic, part black, part white
106	Part "Spanish," part black
107	Part "Spanish," part Indian
108	Part "Spanish," part Asian
121	Mexican and Puerto Rican
122	Mexican and Central or South American
123	Mexican and Cuban
124	Puerto Rican and Central or South American
125	Puerto Rican and Cuban
126	Cuban and Central or South American
127	Mexican and Jamaican

128	Puerto Rican and Jamaican
129	Central or South American and Jamaican
130	Cuban and Jamaican
131	Dominican and Mexican
132	Dominican and Puerto Rican
133	Dominican and Central or South American
134	Dominican and Cuban
135	Mexican and European
136	Puerto Rico and European
137	Central or South American and European
138	Cuban and European
139	Trinidad and Mexican
140	Trinidad and Puerto Rican
141	Trinidad and Central or South American
142	Trinidad and Cuban
143	Mexican and Asian
144	Puerto Rican and Asian
145	Central or South American and Asian
201	Biracial (nonspecific)
202	White and black
203	White and American Indian
204	White and Native Hawaiian
205	White and Other Pacific Islander
206	White and Chinese
207	White and Filipino
208	White and Japanese
209	White and Asian Indian
210	White and Korean
211	White and Vietnamese
212	White and Other Asian
213	White and Asian (nonspecific)
223	Black and American Indian
224	Black and Native Hawaiian
225	Black and Other Pacific Islander
226	Black and Chinese
227	Black and Filipino
228	Black and Japanese
229	Black and Asian Indian
230	Black and Korean
231	Black and Vietnamese
232	Black and Other Asian
233	Black and Asian (nonspecific)
244	American Indian and Native Hawaiian
245	American Indian and Other Pacific Islander
246	American Indian and Chinese
247	American Indian and Filipino
248	American Indian and Japanese
249	American Indian and Asian Indian
250	American Indian and Korean
251	American Indian and Vietnamese
252	American Indian and Other Asian
253	American Indian and Asian (nonspecific)
265	Native Hawaiian and Other Pacific Islander
266	Native Hawaiian and Chinese
267	Native Hawaiian and Filipino
268	Native Hawaiian and Japanese
269	Native Hawaiian and Asian Indian
270	Native Hawaiian and Korean
271	Native Hawaiian and Vietnamese
272	Native Hawaiian and Other Asian
273	Native Hawaiian and Asian (nonspecific)
286	Other Pacific Islander and Chinese
287	Other Pacific Islander and Filipino
288	Other Pacific Islander and Japanese
289	Other Pacific Islander and Asian Indian

290	Other Pacific Islander and Korean
291	Other Pacific Islander and Vietnamese
292	Other Pacific Islander and Other Asian
293	Other Pacific Islander and Asian (nonspecific)
307	Chinese and Filipino
308	Chinese and Japanese
309	Chinese and Asian Indian
310	Chinese and Korean
311	Chinese and Vietnamese
312	Chinese and Other Asian
328	Filipino and Japanese
329	Filipino and Asian Indian
330	Filipino and Korean
331	Filipino and Vietnamese
332	Filipino and Other Asian
349	Japanese and Asian Indian
350	Japanese and Korean
351	Japanese and Vietnamese
352	Japanese and Other Asian
360	Asian Indian and Korean
361	Asian Indian and Vietnamese
362	Asian Indian and Other Asian
371	Korean and Vietnamese
372	Korean and Other Asian
382	Vietnamese and Other Asian
401	White, black, American Indian
402	White, black, Native Hawaiian
403	White, black, Other Pacific Islander
404	White, black, Chinese
405	White, black, Filipino
406	White, black, Japanese
407	White, black, Asian Indian
408	White, black, Korean
409	White, black, Vietnamese
410	White, black, Other Asian
411	White, black, Asian (nonspecific)
420	White, black, Hispanic
421	White, American Indian, Hispanic
900	Mixed/Mezclado
985	Bad data

E.2 Hispanicity

As with the race questions, Hispanic respondents[3] had the opportunity to specify a Hispanic-origin group by responding "other" to the Hispanic-origin group question (QD04). Also, unlike in the PAPI questionnaire (1999 and prior years), respondents were permitted to select multiple Hispanic-origin groups in response to QD04. However, unlike with the CAI race questions, there was no follow-up question asking a respondents to choose a single group from among multiple groups chosen. The CAI Hispanic-origin group question is as follows.

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

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican

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

3 Central or South American
4 Cuban/Cuban American
5 Other (Specify)

E.2.1 Hispanic-Origin Group Alpha Responses

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

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

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

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

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

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

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

E.2.2 Hispanic-Origin Group Dictionary Codes

Codes were assigned to respondents based either on their response to the first 4 categories of QD05 (codes 1 to 4), or on their Hispanicity alpha-specify responses (codes 11 to 85). Codes 11 to 14 are equivalent to codes 1 to 4, except that the race identification was obtained from the alpha-specify responses. The values of EDQD04 were obtained using these codes (see **Section E.1.2**), which are presented below:

1	Mexican/Mexican American/Mexicano/Chicano
2	Puerto Rican
3	Central or South American
4	Cuban/Cuban American
11	Mexican/Mexican American/Mexicano/Chicano
12	Puerto Rican
13	Central or South American
14	Cuban/Cuban American
21	Mexican/Puerto Rican
22	Mexican/Central or South American
23	Mexican/Cuban
24	Puerto Rican/Central or South American
25	Puerto Rican/Cuban
26	Central or South American/Cuban
31	Hispanic Caribbean (includes Dominican Republic, Santo Domingo)
32	Belize (formerly British Honduras)
33	Dominican (Dominica vs. Dominican Republic unclear)
34	Other Caribbean, possibly Hispanic
35	Portugal/European Spanish/Basque/Canary/Cape Verde
36	"Spanish," non-European versus European unclear
37	Philippines/Guam
38	Spanish Filipino or Spanish Guamanian
50	(All) Hispanic, white, no other information
51	(All) Hispanic, black, no other information
52	(All) Hispanic, American Indian, no other information
53	(All) Hispanic, Asian, no other information
54	(All) Hispanic, no other information
60	Part Hispanic, part white
61	Part Hispanic, part black
62	Part Hispanic, part American Indian
63	Part Hispanic, part Asian
64	Part Hispanic, part black, part white
65	Part "Spanish," part black
66	Part "Spanish," part Indian
67	Part "Spanish," part Asian
70	Other possibly Hispanic (white)
71	Other possibly Hispanic (black)
72	Other possibly Hispanic (American Indian)
73	Other possibly Hispanic (Asian)
74	Other possibly Hispanic (multiracial)
75	Other possibly Hispanic (New Mexico)

76	Other possibly Hispanic (Texas)	80	Other definitely not Hispanic (includes Dominica)
77	Other possibly Hispanic (California)	85	Bad Data / "Mixed" / "Mezclado"

Appendix F

Employment Status Alpha Codes for CAI

This page intentionally left blank.

Appendix F

Employment Status Alpha Codes for CAI

As discussed in **Chapter 5**, the employment status questions in the 1999 computer-assisted interviewing (CAI) questionnaire of the National Household Survey on Drug Abuse (NHSDA) were quite different from the paper-and-pencil interviewing (PAPI) questionnaire used in 1999 and in prior NHSDAs. The questions appeared in the noncore section of the 1999 CAI questionnaire, and a respondent's current employment status was determined from responses to a series of questions (QD26 to QD38) regarding work patterns in the past week and past 12 months.

F.1 Questions Regarding Respondents' Reasons for Not Working in the Past Week

As part of this series of questions, respondents who indicated that they did not work in the week prior to the interview were asked their reason for not working. Respondents who indicated that although they did not work in the past week, but did have a job or business, were routed to question QD30, which asked why they did not work during that week. Respondents who indicated that they did not work in the past week and did not have a job or business were routed to QD31a, which asked why they did not have a job or business during that week. Both of these questions had "Some Other Reason" as a possible response, and respondents who chose this answer were asked to specify the reason. Questions QD30 and QD31a are listed below.

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

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

QD31a: Please look at this card and tell me which one of these reasons **best** describes why you did not have a job or business last week. Just give me the number.

1 Unemployed or on layoff and looking for work
2 On layoff and **not** looking for work
3 Keeping house full time
4 Going to school/training
5 Retired
6 Disabled for work
7 Some other reason

F.2 Not Working Alpha Responses and Dictionary Codes

If the response given to the interviewer to either of these questions was a "7," the interviewer could type in the respondent's answer(s) that did not match the first six choices. To map these responses to the edited variable JOBSTAT, responses were coded using employment status dictionaries, one dictionary for each question. QD30A was the other-specify question for QD30. Individuals were routed there if they said that they had a job or business, and the reason they did not work in the past week could not be described by one of the six choices. The codes for QD30A (21 to 56) follow:

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

QD31B was the other-specify question for QD31A. Individuals were routed there if they said that they did not have a job or business or if they did not answer the question regarding whether they had a job or business, and the reason they did not have a job or business could not be described by one of the six choices. The codes for QD31B (21 to 79) follow:

21 Unemployed or on layoff and looking for work
22 On layoff and not looking for work
23 Keeping house full time
24 Going to school/training
25 Retired
26 Disabled for work
41 Respondent has a full-time job

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

Based on responses to the employment status questions, including the alpha-specify responses provided in QD30a and QD31b, a logically edited employment status variable (JOBSTAT) was created and used as a base variable for creating the final imputed employment status variable EMPSTAT3 (see **Chapter 5** and editing documentation). JOBSTAT had many more categories than the final variable EMPSTAT3. Respondents' JOBSTAT values were recoded into the categories of employment status in the final variable as follows. Note that all respondents aged 12 to 17 were assigned to a single category of the final employment status variable.

1. The following JOBSTAT categories were classified as "Employed full-time": worked full-time last week, work full-time during school year, has full-time job and reason for not working unknown. Furthermore, respondents who indicated that they had a job, but were out during the last week, **and** that they usually work 35 or more hours per week were classified as "Employed full-time" if they had one of the following JOBSTAT

values: has job but out (vacation/sick/temporary absence), has job but out (waiting to report to new job), has job but out (self-employed, no business), has job but out (in school/training), not scheduled/temporary/on-call worker, babysitter, has job and did not want to work last week, has a job during school year (no further information), has a job (no further information).

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

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

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

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

Appendix G

Model Summaries

This page intentionally left blank.

Exhibit G.1 Cigarettes: 12 to 17 Year Olds

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

Exhibit G.2 Cigarettes: 18 to 25 Year Olds

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

Exhibit G.3 Cigarettes: 26+ Year Olds

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

Exhibit G.4 Cigars: 12 to 17 Year Olds

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

Exhibit G.5 Cigars: 18 to 25 Year Olds

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

Exhibit G.6 Cigars: 26+ Year Olds

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

Exhibit G.7 Pipes: 12 to 17 Year Olds

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

Exhibit G.8 Pipes: 18 to 25 Year Olds

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

Exhibit G.9 Pipes: 26+ Year Olds

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

Exhibit G.10 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	Race; Gender; Census Region; MSA; Cigarette Lifetime Indicator	Age; Race; Gender; Age^2; Age^3; Cigarette Lifetime Indicator; Age*Race; Age*Gender; Gender*Race; Age^2*Race; Age^2*Gender; Age^3*Race; Age^3*Gender; MSA; State Rank
Recency	Smokeless Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin	Smokeless Tobacco: Age; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Chewing Tobacco: Age; Age^2; Age^3; Race; Gender; Gender*Race; Age*Gender; Age*Race; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin Snuff: Age; Age^2; Age^3; Race; Gender; Census Region; MSA; State Rank; Imputation-Revised Cigarette Recency; Lifetime Indicators of Cigars, Pipes, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, and Heroin
12-Month Frequency	N/A	N/A

Exhibit G.10 (continued)

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

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

Exhibit G.11 (continued)

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

Exhibit G.12 Smokeless Tobacco (Chewing Tobacco and Snuff): 26+ Year Olds

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

Exhibit G.12 (continued)

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

Exhibit G.13 Alcohol: 12 to 17 Year Olds

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

Exhibit G.13 (continued)

Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Alcohol 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, and Cigars Age at First Use

Exhibit G.14 Alcohol: 18 to 25 Year Olds

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

Exhibit G.14 (continued)

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

Exhibit G.15 Alcohol: 26+ Year Olds

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

Exhibit G.15 (continued)

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

Exhibit G.16 Inhalants: 12 to 17 Year Olds

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

Exhibit G.16 (continued)

Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Inhalants 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol and Cigars Age at First Use

Exhibit G.17 Inhalants: 18 to 25 Year Olds

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

Exhibit G.17 (continued)

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

Exhibit G.18 Inhalants: 26+ Year Olds

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

Exhibit G.19 Marijuana: 12 to 17 Year Olds

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

Exhibit G.19 (continued)

Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Marijuana 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants and Cigars Age at First Use

Exhibit G.20 Marijuana: 18 to 25 Year Olds

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

Exhibit G.20 (continued)

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

Exhibit G.21 Marijuana: 26+ Year Olds

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

Exhibit G.21 (continued)

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

Exhibit G.22 Hallucinogens: 12 to 17 Year Olds

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

Exhibit G.22 (continued)

Age at First Use	Race; Gender; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives	Age; Gender; Race; State Rank; Age^2; Age^3; Age*Race; Gender*Race; Age*Gender; MSA; Census Region; Imputation-Revised Cigarettes, Cigars, Smokeless Tobacco, Pipes, Alcohol, Marijuana, Inhalants, Cocaine, Crack, Heroin, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, and Sedatives; Imputation-Revised Hallucinogens 12-Month and 30-Day Frequency; Imputation-Revised Cigarette, Cigarette Daily, Smokeless Tobacco, Alcohol, Inhalants, Marijuana and Cigars Age at First Use

Exhibit G.23 Hallucinogens: 18 to 25 Year Olds

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

Exhibit G.23 (continued)

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

Exhibit G.24 Hallucinogens: 26+ Year Olds

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

Exhibit G.25 Pain Relievers: 12 to 17 Year Olds

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

Exhibit G.26 Pain Relievers: 18 to 25 Year Olds

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

Exhibit G.27 Pain Relievers: 26+ Year Olds

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

Exhibit G.28 Tranquilizers: 12 to 17 Year Olds

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

Exhibit G.29 Tranquilizers: 18 to 25 Year Olds

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

Exhibit G.30 Tranquilizers: 26+ Year Olds

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

Exhibit G.31 Stimulants: 12 to 17 Year Olds

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

Exhibit G.32 Stimulants: 18 to 25 Year Olds

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

Exhibit G.33 Stimulants: 26+ Year Olds

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

Exhibit G.34 Sedatives: 12 to 17 Year Olds

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

Exhibit G.35 Sedatives: 18 to 25 Year Olds

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

Exhibit G.36 Sedatives: 26+ Year Olds

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

Exhibit G.37 Cocaine: 12 to 17 Year Olds

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

Exhibit G.37 (continued)

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

Exhibit G.38 Cocaine: 18 to 25 Year Olds

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

Exhibit G.38 (continued)

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

Exhibit G.39 Cocaine: 26+ Year Olds

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

Exhibit G.39 (continued)

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

Exhibit G.40 Heroin: 12 to 17 Year Olds

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

Exhibit G.41 Heroin: 18 to 25 Year Olds

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

Exhibit G.42 Heroin: 26+ Year Olds

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

Roster Imputations

Exhibit G.43 12 to 17 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total people in household (Screener)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener)
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size
Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household	Age; Age^2; Gender; Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household
Other family present in Household (IRFAMSKP)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household

Exhibit G.44 18 to 25 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Greater Than 64 Years Old in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Marital Status; Employment Status; Education
Other family present in Household (IRFAMSKP)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years old in household; Marital Status; Employment Status; Education

Exhibit G.45 26 to 64 Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Greater Than 64 in Household (IRHH65)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Marital Status; Employment Status; Education
Other Family Present in Household (IRFAMSKP)	Age; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education

Exhibit G.46 65+ Year Olds

	Variables Included in Response Propensity	Variables Included in Roster Model
Imputation-Revised Household Size (IRHHSIZE)	Age; Gender; Race; Census Region; MSA; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status;	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Total People in Household (Screener); Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Younger Than 18 Years Old in Household (IRKID17)	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Number of Eligible 12 to 17 in household (Screener); Imputation-Revised household size Marital Status; Employment Status; Education
Imputation-Revised Number of Persons Greater Than 64 Years old in Household (IRHH65)	Age; Gender; Race; Census Region; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Marital Status; Education	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years old in household; Marital Status; Employment Status; Education
Other Family Present in Household (IRFAMSKP)	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Region; MSA; Percent Hispanic in Segment; Percent Owner Occupied in Segment; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years old in household; Marital Status; Employment Status; Education	Age; Gender; Race; Census Region; Imputation-Revised household size; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Imputation-Revised Number of Persons Greater Than 64 Years Old in Household; Marital Status; Employment Status; Education

Income Imputations

Exhibit G.47 Response Propensity for Dichotomous Income Indicators

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12 to 17	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank
18 to 25	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status
26 to 64	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status
65+	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status

Exhibit G.48 Modeling for Dichotomous Income Indicators: 12 to 17 Year Olds

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank;
Supplemental Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State , and Intermediate Family Social Security
Welfare Payments	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security
Welfare Services	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments
Investment Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services
Child Support	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income

Exhibit G.48 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
Wages	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support
Other Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage
Food Stamps	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income
# Welfare Months	Age; Gender; Race; Census Region; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Investment Income; Intermediate Family Wage
Total Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Intermediate Family Food Stamps

Exhibit G.49 Modeling for Dichotomous Income Indicators: 18 to 25 Year Olds

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status
Supplemental Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Marital Status; Education; Employment Status
Welfare Payments	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Marital Status; Education; Employment Status
Welfare Services	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Marital Status; Education; Employment Status
Investment Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Marital Status; Education; Employment Status
Child Support	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Marital Status; Education; Employment Status

Exhibit G.49 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
Wages	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Marital Status; Education; Employment Status
Other Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage
Food Stamps	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Marital Status; Education; Employment Status
# Welfare Months	Age; Gender; Race; Census Region; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Investment Income; Intermediate Family Wage; Marital Status; Education; Employment Status
Total Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education; Employment Status

Exhibit G.50 Modeling for Dichotomous Income Indicators: 26 to 64 Year Olds

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status
Supplemental Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Marital Status; Education; Employment Status
Welfare Payments	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Marital Status; Education; Employment Status
Welfare Services	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Marital Status; Education; Employment Status
Investment Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Marital Status; Education; Employment Status
Child Support	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Marital Status; Education; Employment Status

Exhibit G.50 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
Wages	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Marital Status; Education; Employment Status
Other Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage
Food Stamps	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Marital Status; Education; Employment Status
# Welfare Months	Age; Gender; Race; Age^2; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA;Number of Adults in Household; Intermediate Family Welfare Services; Intermediate Family Wage; Intermediate Family Other Income; Marital Status; Education; Employment Status
Total Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education; Employment Status

Exhibit G.51 Modeling for Dichotomous Income Indicators: 65+ Year Olds

	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status
Supplemental Security	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Marital Status; Education; Employment Status
Welfare Payments	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Marital Status; Education; Employment Status
Welfare Services	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Marital Status; Education; Employment Status
Investment Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Marital Status; Education; Employment Status
Child Support	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Marital Status; Education; Employment Status

Exhibit G.51 (continued)

	Variables Included in Income Model (Dichotomous Income Indicators)
Wages	Age; Gender; Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Marital Status; Education; Employment Status
Other Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage
Food Stamps	Age; Gender; Race; Age^2; Gender*Age; Gender*Age^2; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Marital Status; Education; Employment Status
# Welfare Months	Age; Gender; Race; Gender*Race; Age^2; Gender*Age; Gender*Age^2; MSA; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Welfare Payments; Intermediate Family Wage; Intermediate Family Food Stamps; Marital Status;
Total Income	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Intermediate Family Social Security; Intermediate Family Supplemental Security; Intermediate Family Welfare Payments; Intermediate Family Welfare Services; Intermediate Family Investment Income; Intermediate Family Child Support; Intermediate Family Wage; Intermediate Family Other Income; Intermediate Family Food Stamps; Marital Status; Education; Employment Status

Exhibit G.52 Response Propensity for Income Finer Categories

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
12 to 17	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
26 to 64	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
65+	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)

Exhibit G.53 Modeling for Income Finer Categories

Age Group	Variables Included in Income Models (Finer Categorization)
12 to 17	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
18 to 25	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
26 to 64	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Age*Race; Gender*Age^2; Age^2* Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)
65+	Age; Gender; Race; Gender*Race; Gender*Age; Age*Race; Census Region; MSA; Percent Hispanic in Segment; Percent Black in Segment; Percent Owner Occupied in Segment; Number of Adults in Household; Imputation-Revised Number of Persons Greater than 64 Years Old in Household; Imputation-Revised Number of Persons Younger Than 18 Years Old in Household; Income State Rank; Marital Status; Education; Employment Status; Imputation-Revised Family Social Security; Imputation-Revised Family Supplemental Security; Imputation-Revised Family Welfare Payments; Imputation-Revised Family Welfare Services; Imputation-Revised Family Investment Income; Imputation-Revised Family Child Support; Imputation-Revised Family Wages; Imputation-Revised Family Other Income; Imputation-Revised Family Food Stamps; Imputation-Revised Family Income (Dichotomous)

Health Insurance Imputations

Exhibit G.54 12 to 17 Year Olds

	Variables Included in Response Propensity	Variables Included in Health Insurance Model
Imputation-Revised Overall Health Insurance (IRINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size
Imputation-Revised Private Health Insurance (IRPINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size

Exhibit G.55 18 to 25 Year Olds

	Variables Included in Response Propensity	Variables Included in Health Insurance Model
Imputation-Revised Overall Health Insurance (IRINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size
Imputation-Revised Private Health Insurance (IRPINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size

Exhibit G.56 26 to 64 Year Olds

	Variables Included in Response Propensity	Variables Included in Health Insurance Model
Imputation-Revised Overall Health Insurance (IRINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size
Imputation-Revised Private Health Insurance (IRPINSUR)	Age; Age^2; Gender; Race; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Age^2; Gender*Race; Gender*Age; Gender*Age^2; Age*Race; Age^2*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size

Exhibit G.57 65+ Year Olds

	Variables Included in Response Propensity	Variables Included in Health Insurance Model
Imputation-Revised Overall Health Insurance (IRINSUR)	Gender; Race; Gender*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; MSA; Marital Status; Education Status; Employment Status; Percent Hispanic in Segment; Percent Black in Segment; Household Size
Imputation-Revised Private Health Insurance (IRPINSUR)	Gender; Race; Gender*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size	Age; Gender; Race; Gender*Race; MSA; Marital Status; Education Status; Employment Status; Percent Owner Occupied in Segment; Percent Hispanic in Segment; Percent Black in Segment; Household Size

Appendix H

Error in First Run of Multivariate Recency and Frequency of Use Imputations

Appendix H

Error in First Run of Multivariate Recency and Frequency of Use Imputations

H.1 Introduction

A programming error in the first run of the 1999 National Household Survey on Drug Abuse (NHSDA) imputations of recency and frequency of use was discovered, requiring the reimputation of these two variables as well as age at first use. For most drugs, the error did not have a noticeable effect. In short, the error was limited to those cases in which recency was not completely known, where it was necessary to maintain consistency between the 30-day and 12-month frequency-of-use variables. Because the tobacco variables (cigarettes, cigars, pipes, and smokeless tobacco) and the pill variables (pain relievers, stimulants, tranquilizers, and sedatives) do not have both the 30-day and 12-month frequency-of-use variables, they were not affected. Moreover, the modules where values for more than one substance corresponded to a single predictive mean vector (cocaine and crack; hallucinogens, LSD, and PCP; and stimulants and methamphetamines) were handled differently, so that the error did not occur with the imputation of missing values for those drugs. Hence, the error was limited to the imputations in the alcohol, marijuana, inhalants, and heroin variables, with the greatest effect concentrated in alcohol, marijuana, and inhalants. The error resulted in an overestimate in the prevalence of past year and past month usage of these four substances. Due to minor adjustments in the methodology between the first time the imputations were done and the discovery of the error[1], it was decided (for the sake of consistency) to redo the imputations for recency of use, frequency of use, and age at first use for all drugs. Because of this and the fact that the imputation procedure is not deterministic, the most recently published drug use estimates differ slightly from those first published for the 1999 survey.

H.2 How the Error Was Discovered

Although the imputation of recency and frequency of use is multivariate, the drug use measures were modeled sequentially with recency of use modeled first, followed by the

[1]In the first run of the drug use imputations, the primary focus of the imputation procedures was to find a neighborhood with 30 candidate donors. Subsequent theoretical considerations supported the notion that the predicted mean of the donor and recipient had to be as close as possible (within 5%), whereby if no donors were available within 5% of the recipient's predicted mean, the closest donor was chosen and a neighborhood was not used. The 1999 drug use imputations were redone using this criterion.

12-month and 30-day frequency-of-use variables. Hence, univariate provisional imputations were required so that provisional recency-of-use values could be used in the 12-month frequency-of-use imputations, provisional 12-month frequency-of-use imputed values could be used in the 30-day frequency-of-use imputations, and, for alcohol, provisional 30-day frequency-of-use imputed values could be used in the 30-day binge drinking imputations. In these provisional imputations, no attempt was made to maintain consistency between the 30-day and 12-month frequencies. Hence, the univariate imputations were unaffected by the error.

New quality control checks were instituted on the 2000 drug imputations, which were also applied to the 1999 imputed data. For each variable, the distribution of the univariate imputed values was compared to the distribution of the multivariate imputed values. There was a large discrepancy between the distributions of recency of use for the univariate and multivariate imputed values, with a large concentration of past month users in the multivariate imputed values distribution. This discrepancy was too large to be explained by the difference between the univariate and multivariate imputations.

H.3 Description of the Error

If a respondent is a past month user of a substance, he or she would have values for all the frequency-of-use variables. However, if the respondent is a user in the past year but not the past month, he or she would not have values for the 30-day frequency-of-use variables (30-day frequency of use and 30-day binge drinking frequency). Moreover, for respondents who did not use in the past year, they would not have values for any of the frequency-of-use variables. The editing team assigns skip codes for these respondents: a "93" for the 30-day frequency variables and a "993" for the 12-month frequency variables. For the 12-month frequency, the variable that is actually used in the imputation of missing values is the proportion of a period within the past year that the donor used. This "period within the past year" is the maximum period within the past year that the donor could have used, which is either the full year or part of the past year, depending upon the donor's month of first use. For potential donors who are not past year users, this proportion would be missing.

The donor pool for respondents whose recency is not completely known would consist of respondents with a variety of values for recency and frequency of use, including skip codes for frequency of use where applicable. For example, if a respondent is a lifetime user of marijuana, the following are possible donors: past month user with valid values for 12-month and 30-day frequencies; past year but not past month user with valid values for 12-month frequency of use, and the skip code for 30-day frequency of use; and a lifetime but not past year user with skip codes for 12-month and 30-day frequencies, and missing values for the proportion of the past year or part year that the donor used.

The assigned 12-month frequency of use variable is a product of the donor’s proportion of the past year or part year that the donor used and the recipient's maximum period in the past year he or she could have used. This product must be greater than the donor’s 30-day frequency, provided the donor is a past month user. The error resulted from implementing this consistency constraint without checking that the donor was a past month user. Hence, this affected cases where the nonspecific recency of use was "past year use" or "lifetime use." If a potential donor was a past year but not a past month user, he or she would be excluded from the donor pool if his or her 12-month frequency product (defined above) was less than 93, the skip code for 30-day frequency of use. In those cases where no information was available on recency of use (lifetime users), a potential donor who was a lifetime but not past year user would not ever be considered as a donor. This was because the proportion of the past year or part year that the donor used was missing for these cases, so that his or her 12-month frequency product would also be missing. Since SAS was used for analyzing these data and the missing code in SAS is considered "smaller" than any other value, the product that represents the donated 12-month frequency would always be less than the donor's value for the 30-day frequency, 93.

H.4 How the Error Was Corrected

In the revised programs for the multivariate imputation of recency and frequency of use, the consistency constraints that are applied depend upon the recency of use of the potential donor. Hence, donors who are past month users have one set of consistency constraints applied, past year but not past month users have another set, and lifetime but not past year users have yet another set. With the revised imputed values, no obvious difference was apparent when comparing the distribution of univariate provisional imputed values with the distribution of multivariate final imputed values.

The effect of the imputation error on the drug use estimates is summarized in **Exhibit H-1**, which follows.

Exhibit H-1. Recency Distribution With and Without the Imputation Error - Total

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Cigarettes	1	17,900	26.83	25.76	17,890	26.82	25.76
	2	4,537	6.80	4.38	4,518	6.77	4.37
	3	4,601	6.90	4.59	4,627	6.94	4.57
	4	11,564	17.34	33.43	11,567	17.34	33.45
	9	28,104	42.13	31.84	28,104	42.13	31.84
Cigars	1	4,783	7.17	5.48	4,781	7.17	5.49
	2	577	8.65	6.27	5,790	8.68	6.29
	3	5,194	7.79	6.08	5,171	7.75	6.06
	4	5708	8.56	17.72	5,714	8.57	17.72
	9	45,250	67.84	64.45	45,250	67.84	64.45
Smokeless Tobacco	1	2,756	4.13	3.42	2,761	4.14	3.42
	2	1,589	2.38	1.24	1,576	2.36	1.23
	3	2,036	3.05	1.79	2,044	3.06	1.77
	4	6,408	9.61	12.64	6,408	9.61	12.67
	9	53,917	80.83	80.91	53,917	80.83	80.91
Alcohol	1	27,491	41.21	47.31	26,705	40.03	46.45
	2	10,856	16.27	15.26	11,394	17.08	15.84
	3	7,876	11.81	18.74	8,124	12.18	19.02
	9	20,483	30.71	18.69	20,483	30.71	18.69
Marijuana	1	5,787	8.68	5.05	5,437	8.15	4.73
	2	4,382	6.57	3.80	4,525	6.78	3.91
	3	12,492	18.73	25.71	12,699	19.04	25.92
	9	44,045	66.03	65.44	44,045	66.03	65.44
Inhalants	1	787	1.18	0.46	448	0.67	0.26
	2	1,060	1.59	0.64	1,118	1.68	0.62
	3	5,103	7.65	6.65	5,384	8.07	6.87
	9	59,756	89.58	92.25	59,756	89.58	92.25
Cocaine	1	597	0.90	0.68	592	0.89	0.70
	2	1,124	1.69	0.99	1,135	1.70	0.99
	3	4,180	6.27	9.82	4,174	6.26	9.80
	9	60,805	91.15	88.51	60,805	91.15	88.51

Exhibit H-1. (continued)

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Crack	1	130	0.19	0.19	130	0.19	0.19
	2	273	0.41	0.28	274	0.41	0.28
	3	1,128	1.69	2.20	1,127	1.69	2.20
	9	65,175	97.70	97.33	65,175	97.70	97.33
Heroin	1	98	0.15	0.09	70	0.10	0.07
	2	93	0.14	0.09	99	0.15	0.09
	3	534	0.80	1.20	556	0.83	1.22
	9	65,981	98.91	98.62	65,981	98.91	98.62
Hallucino-gens	1	667	1.00	0.41	681	1.02	0.42
	2	1,757	2.63	1.02	1,755	2.63	1.03
	3	5,569	8.35	9.90	5,557	8.33	9.89
	9	58,713	88.02	88.67	58,713	88.02	88.67
Stimulants	1	470	0.70	0.42	470	0.70	0.43
	2	842	1.26	0.62	832	1.25	0.61
	3	3,224	4.83	6.16	3,234	4.85	6.16
	9	62,170	93.20	92.80	62,170	93.20	92.80
Sedatives	1	112	0.17	0.11	110	0.16	0.10
	2	191	0.29	0.18	189	0.28	0.18
	3	1,068	1.60	3.22	1,072	1.61	3.22
	9	65,335	97.94	96.50	65,335	97.94	96.50
Pain Relievers	1	1,323	1.98	1.20	1,322	1.98	1.19
	2	2,089	3.13	1.80	2,105	3.16	1.79
	3	3,677	5.51	5.99	3,662	5.49	6.02
	9	59,617	89.37	91.01	59,617	89.37	91.01
Tranquil-izers	1	422	0.63	0.50	438	0.66	0.50
	2	820	1.23	0.76	807	1.21	0.74
	3	2,346	3.52	5.01	2,343	3.51	5.03
	9	63,118	94.62	93.73	63,118	94.62	93.73

[1]The recency levels are defined as: (1) past month user; (2) past year but not past month user; (3) non-tobacco substances: lifetime but not past year user; tobacco only; past 3 years but not past year user; (4) tobacco only; lifetime but not past 3 years user; (9) nonuser.

Exhibit H-2. Recency Distribution With and Without the Imputation Error - 12 to 17

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Cigarettes	1	3,826	15.09	14.89	3,821	15.07	14.90
	2	2,092	8.25	8.59	2,074	8.18	8.50
	3	2,070	8.16	8.31	2,100	8.28	8.44
	4	1,385	5.46	5.29	1,378	5.43	5.24
	9	15,984	63.04	62.93	15,984	63.04	62.93
Cigars	1	1,349	5.32	5.36	1,349	5.32	5.36
	2	1,798	7.09	7.19	1,799	7.09	7.21
	3	1,321	5.21	5.43	1,311	5.17	5.38
	4	401	1.58	1.58	410	1.62	1.62
	9	20,488	80.80	80.43	20,488	80.80	80.43
Smokeless Tobacco	1	591	2.33	2.28	594	2.34	2.29
	2	650	2.56	2.37	637	2.51	2.32
	3	754	2.97	2.96	763	3.01	2.99
	4	584	2.30	2.18	585	2.31	2.19
	9	22,778	89.83	90.21	22,778	89.83	90.21
Alcohol	1	4,782	18.86	18.64	4278	16.87	16.54
	2	4,037	15.92	16.23	4366	17.22	17.59
	3	2,018	7.96	8.00	2193	8.65	8.74
	9	14,520	57.26	57.13	14,520	57.26	57.13
Marijuana	1	1,999	7.88	7.75	1,859	7.33	7.22
	2	1,665	6.57	6.68	1,732	6.83	6.93
	3	1,132	4.46	4.30	1,205	4.75	4.57
	9	20,561	81.09	81.27	20,561	81.09	81.27
Inhalants	1	532	2.10	1.90	301	1.19	1.05
	2	651	2.57	2.70	704	2.78	2.86
	3	1,203	4.74	4.53	1,381	5.45	5.22
	9	22,971	90.59	90.87	22,971	90.59	90.87
Cocaine	1	135	0.53	0.53	137	0.54	0.53
	2	251	0.99	1.03	260	1.03	1.07
	3	214	0.84	0.81	203	0.80	0.77
	9	24,757	97.63	97.62	24,757	97.63	97.62

Exhibit H-2. (continued)

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Crack	1	25	0.10	0.08	26	0.10	0.08
	2	73	0.29	0.31	75	0.30	0.31
	3	63	0.25	0.23	60	0.24	0.23
	9	25,196	99.37	99.38	25,196	99.37	99.38
Heroin	1	40	0.16	0.18	33	0.13	0.17
	2	29	0.11	0.12	30	0.12	0.12
	3	36	0.14	0.13	42	0.17	0.16
	9	25,252	99.59	99.56	25,252	99.59	99.56
Hallucino-gens	1	267	1.05	1.10	272	1.07	1.10
	2	681	2.69	2.69	684	2.70	2.74
	3	506	2.00	1.88	498	1.96	1.84
	9	23,903	94.27	94.33	23,903	94.27	94.33
Stimulants	1	182	0.72	0.67	181	0.71	0.65
	2	374	1.47	1.44	372	1.47	1.47
	3	475	1.87	1.83	478	1.89	1.82
	9	24,326	95.93	96.05	24,326	95.93	96.05
Sedatives	1	51	0.20	0.20	51	0.20	0.19
	2	80	0.32	0.29	80	0.32	0.29
	3	88	0.35	0.33	88	0.35	0.34
	9	25,138	99.14	99.18	25,138	99.14	99.18
Pain Relievers	1	560	2.21	2.16	564	2.22	2.14
	2	795	3.14	3.18	816	3.22	3.32
	3	716	2.82	2.82	691	2.73	2.70
	9	23,286	91.83	91.84	23,286	91.83	91.84
Tranquil-izers	1	122	0.48	0.48	132	0.52	0.51
	2	256	1.01	1.06	255	1.01	1.05
	3	230	0.91	0.93	221	0.87	0.91
	9	24,749	97.60	97.53	24,749	97.60	97.53

[1]The recency levels are defined as: (1) past month user; (2) past year but not past month user; (3) non-tobacco substances: lifetime but not past year user; tobacco only; past 3 years but not past year user; (4) tobacco only; lifetime but not past 3 years user; (9) nonuser.

Exhibit H-3. Recency Distribution With and Without the Imputation Error - 18 to 25

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Cigarettes	1	8,722	39.77	39.68	8,718	39.75	39.67
	2	1,733	7.90	7.80	1,733	7.90	7.83
	3	1,732	7.90	7.80	1,731	7.89	7.78
	4	3,010	13.72	13.63	3,015	13.75	13.64
	9	6,736	30.71	31.09	6,736	30.71	31.09
Cigars	1	2,402	10.95	11.54	2,397	10.93	11.50
	2	2,800	12.77	13.37	2,813	12.83	13.45
	3	2,667	12.16	12.28	2,657	12.11	12.26
	4	1,563	7.13	6.75	1,565	7.14	6.73
	9	12,501	57.00	56.06	12,501	57.00	56.06
Smokeless Tobacco	1	1,373	6.26	5.73	1,375	6.27	5.74
	2	752	3.43	3.48	752	3.43	3.47
	3	1,005	4.58	4.41	1,006	4.59	4.41
	4	2,758	12.57	12.19	2,755	12.56	12.19
	9	16,045	73.15	74.19	16,045	73.15	74.19
Alcohol	1	12,604	57.47	58.00	12,443	56.73	57.24
	2	3,766	17.17	16.97	3,889	17.73	17.54
	3	2,061	9.40	8.94	2,099	9.57	9.12
	9	3,502	15.97	16.09	3,502	15.97	16.09
Marijuana	1	3,039	13.86	14.81	2,899	13.22	14.22
	2	2,170	9.89	10.02	2,224	10.14	10.23
	3	4,931	22.48	21.98	5,017	22.87	22.35
	9	11,793	53.77	53.19	11,793	53.77	53.19
Inhalants	1	222	1.01	1.08	127	0.58	0.64
	2	377	1.72	1.91	394	1.80	1.98
	3	2,411	10.99	11.11	2489	11.35	11.48
	9	18,923	86.28	85.90	18,923	86.28	85.90
Cocaine	1	342	1.56	1.73	329	1.50	1.67
	2	711	3.24	3.53	716	3.26	3.57
	3	1,466	6.68	6.62	1,474	6.72	6.64
	9	19,414	88.52	88.12	19,414	88.52	88.12

Exhibit H-3. (continued)

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Crack	1	63	0.29	0.26	61	0.28	0.26
	2	155	0.71	0.70	154	0.70	0.69
	3	506	2.31	2.33	509	2.32	2.33
	9	21,209	96.70	96.71	21,209	96.70	96.71
Heroin	1	44	0.20	0.20	27	0.12	0.13
	2	56	0.26	0.30	61	0.28	0.33
	3	248	1.13	1.25	260	1.19	1.30
	9	21,585	98.41	98.25	21,585	98.41	98.25
Hallucino-gens	1	381	1.74	1.84	388	1.77	1.87
	2	1,021	4.66	4.95	1015	4.63	4.90
	3	2,679	12.21	12.50	2678	12.21	12.51
	9	17,852	81.39	80.71	17,852	81.39	80.71
Stimulants	1	230	1.05	1.03	232	1.06	1.05
	2	389	1.77	1.92	384	1.75	1.90
	3	1,329	6.06	6.04	1,332	6.07	6.04
	9	19,985	91.12	91.01	19,985	91.12	91.01
Sedatives	1	45	0.21	0.21	44	0.20	0.20
	2	85	0.39	0.43	83	0.38	0.41
	3	273	1.24	1.34	276	1.26	1.36
	9	21,530	98.16	98.03	21,530	98.16	98.03
Pain Relievers	1	582	2.65	2.61	578	2.64	2.61
	2	1,016	4.63	4.94	1,011	4.61	4.98
	3	1,603	7.31	7.61	1,612	7.35	7.57
	9	18,732	85.41	84.84	18,732	85.41	84.84
Tranquil-izers	1	213	0.97	1.03	220	1.00	1.06
	2	437	1.99	2.13	427	1.95	2.08
	3	967	4.41	4.71	970	4.42	4.73
	9	20,316	92.63	92.13	20,316	92.63	92.13

[1]The recency levels are defined as: (1) past month user; (2) past year but not past month user; (3) non-tobacco substances: lifetime but not past year user; tobacco only; past 3 years but not past year user; (4) tobacco only; lifetime but not past 3 years user; (9) nonuser.

Exhibit H-4. Recency Distribution With and Without the Imputation Error - 26 or Older

Drug	Level[1]	Data with Imputation Error: Unweighted Count	Data with Imputation Error: Unweighted %	Data with Imputation Error: Weighted %	Data after Imputation Error Fixed: Unweighted Count	Data after Imputation Error Fixed: Unweighted %	Data after Imputation Error Fixed: Weighted %
Cigarettes	1	5,352	27.56	24.91	5,351	27.56	24.91
	2	712	3.67	3.22	711	3.66	3.23
	3	799	4.12	3.54	796	4.10	3.50
	4	7,169	36.92	40.61	7,174	36.95	40.65
	9	5,384	27.73	27.71	5,384	27.73	27.71
Cigars	1	1,032	5.32	4.48	1,035	5.33	4.50
	2	1,173	6.04	4.95	1,178	6.07	4.95
	3	1,206	6.21	5.13	1,203	6.20	5.12
	4	3,744	19.28	21.78	3,739	19.26	21.77
	9	12,261	63.15	63.67	12,261	63.15	63.67
Smokeless Tobacco	1	792	4.08	3.19	792	4.08	3.19
	2	187	0.96	0.71	187	0.96	0.71
	3	277	1.43	1.19	275	1.42	1.16
	4	3,066	15.79	14.15	3,068	15.80	14.18
	9	15,094	77.74	80.77	15,094	77.74	80.77
Alcohol	1	10,105	52.04	49.43	9,984	51.42	48.73
	2	3,053	15.72	14.84	3,139	16.17	15.32
	3	3,797	19.56	21.86	3,832	19.74	22.08
	9	2,461	12.68	13.87	2,461	12.68	13.87
Marijuana	1	749	3.86	3.05	679	3.50	2.79
	2	547	2.82	2.36	569	2.93	2.43
	3	6,429	33.11	29.27	6,477	33.36	29.45
	9	11,691	60.21	65.33	11,691	60.21	65.33
Inhalants	1	33	0.17	0.16	20	0.10	0.09
	2	32	0.16	0.15	20	0.10	0.08
	3	1,489	7.67	6.19	1,514	7.80	6.33
	9	17,862	92.00	93.50	17,862	92.00	93.50
Cocaine	1	120	0.62	0.52	126	0.65	0.56
	2	162	0.83	0.56	159	0.82	0.55
	3	2,500	12.88	11.59	2,497	12.86	11.56
	9	16,634	85.67	87.33	16,634	85.67	87.33

Exhibit H-4. (continued)

Drug	Level[1]	Data with Imputation Error			Data after Imputation Error Fixed		
		Un-weighted Count	Un-weighted %	Weighted %	Un-weighted Count	Un-weighted %	Weighted %
Crack	1	42	0.22	0.19	43	0.22	0.19
	2	45	0.23	0.21	45	0.23	0.21
	3	559	2.88	2.45	558	2.87	2.45
	9	18,770	96.67	97.15	18,770	96.67	97.15
Heroin	1	14	0.07	0.06	10	0.05	0.05
	2	8	0.04	0.05	8	0.04	0.05
	3	250	1.29	1.34	254	1.31	1.35
	9	19,144	98.60	98.55	19,144	98.60	98.55
Hallucino-gens	1	19	0.10	0.08	21	0.11	0.08
	2	55	0.28	0.14	56	0.29	0.14
	3	2,384	12.28	10.56	2,381	12.26	10.55
	9	16,958	87.34	89.23	16,958	87.34	89.23
Stimulants	1	58	0.30	0.29	57	0.29	0.30
	2	79	0.41	0.29	76	0.39	0.27
	3	1,420	7.31	6.77	1,424	7.33	6.78
	9	17,859	91.98	92.65	17,859	91.98	92.66
Sedatives	1	16	0.08	0.08	15	0.08	0.08
	2	26	0.13	0.12	26	0.13	0.13
	3	707	3.64	3.93	708	3.65	3.92
	9	18,667	96.14	95.87	18,667	96.14	95.87
Pain Relievers	1	181	0.93	0.83	180	0.93	0.82
	2	278	1.43	1.09	278	1.43	1.05
	3	1,358	6.99	6.16	1,359	7.00	6.21
	9	17,599	90.64	91.93	17,599	90.64	91.93
Tranquil-izers	1	87	0.45	0.42	86	0.44	0.40
	2	127	0.65	0.49	125	0.64	0.47
	3	1,149	5.92	5.61	1,152	5.93	5.65
	9	18,053	92.98	93.48	18,053	92.98	93.48

[1]The recency levels are defined as: (1) past month user; (2) past year but not past month user; (3) non-tobacco substances: lifetime but not past year user; tobacco only; past 3 years but not past year user; (4) tobacco only; lifetime but not past 3 years user; (9) nonuser.

Appendix I

Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

This page intentionally left blank.

Appendix I

Numbers of Respondents Meeting Likeness Constraints on Sets of Eligible Donors

I. 1 Introduction

For all the variables for which imputations were implemented, whether the predictive mean neighborhood (PMN) was univariate or multivariate, restrictions were placed upon the neighborhood prior to the assignment of imputed values. The pool of potential donors for a given recipient was restricted according to logical and likeness constraints, where the likeness constraints were loosened if donors could not be found, but logical constraints could not be loosened. Since logical constraints (summarized in **Appendix J**) cannot be loosened, the attempt to find a donor under those constraints is either successful or not successful; there is no opportunity to loosen the constraints.[1] Such an opportunity does exist, however, with likeness constraints. If no donors were available under the most stringent set of constraints, constraints were loosened, one at a time, until a donor could be found. This appendix summarizes the number of cases for which donors were available under each of the various likeness constraints, starting with the most stringent constraint.

Although statistical imputation of the drug use or income variables could not proceed separately within each State due to insufficient pools of donors, information about the State of residence of each respondent was incorporated in the PMN procedure. For the drug use variables, in the hot-deck step of PMN, respondents were separated into three State usage-level categories for each drug depending on the response variable of interest. Respondents from States with high usage of a given drug were placed in one category, respondents from medium usage States into another, and the remainder into a third category. The States were separated into three income groups for the income variables, depending upon the proportion of families with incomes greater than or equal to $20,000. As with the drug use variables, respondents from high-income States (by this measure) were placed in one category, respondents from medium income states into another category, and the remainder into a third category. In the exhibits that follow, this variable is identified as the "State rank" for the drug use and income variables. It was used as a likeness constraint, where the set of eligible donors for each recipient was restricted so that donors and recipients were both from States with the same State rank.

[1] Logical constraints define what is normally referred to as an "imputation class."

The phrase "Donor's predicted means each within *x*% 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*% of recipient's predicted mean" appears in each of the univariate imputation exhibits. In either case, it represents one of the likeness constraints. It also defines the neighborhood. Once this constraint is loosened, the neighborhood is abandoned and the candidate with the predicted mean closest to the recipient's, subject to the constraints that are still on the pool of donors, is chosen as the donor.

I. 2 Drug Variables

The imputation of the drug variables was done separately for three age groups: 12 to 17, 18 to 25, and 26 or older. For each of the drugs, a multivariate imputation was done for the recency and frequency variables, and a univariate imputation was done for the age at first use variable. The exhibits in this appendix show the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

I.2.1 Likeness Constraints for Recency and Frequency Imputation, by Drug

Exhibits I.1 to **I.13** present information on the likeness constraints for recency and frequency imputation for the following drugs: tobacco (i.e., cigarettes, cigars, and smokeless tobacco [chewing tobacco and snuff]), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, sedatives, and stimulants), cocaine, and heroin.

Exhibit I.1 Cigarette Recency and Frequency Imputation

	Frequency		
Likeness Constraints	**12-17**	**18-25**	**26+**
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	89	120	427
(A) State rank of donor = State rank of recipient	64	48	69

Exhibit I.2 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% of recipient's predicted means	305	179	69
(A) State rank of donor = State rank of recipient	60	40	40

Exhibit I.3 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 are the same as recipient's recencies (when nonmissing) (C) Donor's predicted means each within 5% of recipient's predicted means	119	67	14
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	13	13	1
(A) State rank of donor = State rank of recipient	128	98	74

Exhibit I.4 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% of recipient's predicted means	948	124	64
(A) State rank of donor = State rank of recipient	511	264	210

Exhibit I.5 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% of recipient's predicted means	67	5	1
(A) State rank of donor = State rank of recipient	319	106	40

Exhibit I.6 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% of recipient's predicted means	10	131	27
(A) State rank of donor = State rank of recipient	465	252	157
None	7	0	0

Exhibit I.7 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 are the same as recipient's recencies (when nonmissing) (C) Donor's predicted means each within 5% of recipient's predicted means	23	46	1
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	12	13	0
(A) State rank of donor = State rank of recipient	207	143	56

Exhibit I.8 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% of recipient's predicted means	155	65	15
(A) State rank of donor = State rank of recipient	276	172	95

Exhibit I.9 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% of recipient's predicted means	6	15	7
(A) State rank of donor = state rank of recipient	84	85	50

Exhibit I.10 Sedatives Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	0	3	1
(A) State rank of donor = State rank of recipient	46	23	23

Exhibit I.11 Stimulants Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's recency for methamphetamines agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5% of recipient's predicted means	31	74	19
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	33	79	5
(A) State rank of donor = State rank of recipient	176	135	111

Exhibit I.12 Cocaine Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's recency for crack agrees with recipient's recency (when nonmissing) (C) Donor's predicted means each within 5% of recipient's predicted means	3	19	2
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	0	1	1
(A) State rank of donor = State rank of recipient	88	131	65

Exhibit I.13 Heroin Recency and Frequency Imputation

Likeness Constraints	Frequency 12-17	Frequency 18-25	Frequency 26+
(A) State rank of donor = State rank of recipient (B) Donor's predicted means each within 5% of recipient's predicted means	1	0	2
(A) State rank of donor = State rank of recipient	11	19	7

I.2.2 Likeness Constraints for Age at First Use Imputation, by Drug

Exhibits I.14 to **I.26** present information on the likeness constraints for age at first use (AFU) imputation for the following drugs: tobacco (i.e., cigarettes, cigars, and smokeless tobacco [chewing tobacco and snuff]), alcohol, inhalants, marijuana, hallucinogens, psychotherapeutics (i.e., analgesics, tranquilizers, sedatives, and stimulants), cocaine, and heroin.

Exhibit I.14 Cigarette Age at First Use Imputation

	Frequency		
Likeness Constraints	**12-17**	**18-25**	**26+**
(A) Age of donor = Age of recipient (B) State rank of donor = State rank of recipient (C) If recipient did not use in the past year, donor must not have used in the past year (D) Donor's predicted mean within 5% of recipient's predicted mean	526	324	435
(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% of recipient's predicted mean	2	0	20
(A) Age of donor = Age of recipient (B) If recipient did not use in the past year, donor must not have used in the past year	1	0	13
(A) Age of donor = Age of recipient	0	0	1
(A) AFU of donor ≤ Age of recipient,* age of donor ≥ age of recipient	0	0	0
(A) AFU of donor ≤ Age of recipient*	0	0	1

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.15 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% of recipient's predicted mean	340	261	346
(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% of recipient's predicted mean	5	3	47
(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	23
(A) Age of donor = Age of recipient	0	0	2

Exhibit I.16 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 chew if the recipient is missing chew AFU, and for snuff if the recipient is missing snuff AFU) (D) Donor's predicted mean within 5% of recipient's predicted mean	187	196	121
(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 chew if the recipient is missing chew AFU, and for snuff if the recipient is missing snuff AFU) (C) Donor's predicted mean within 5% of recipient's predicted mean	35	6	35
(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 chew if the recipient is missing chew AFU, and for snuff if the recipient is missing snuff AFU)	16	5	51
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,[*] age of donor ≥ age of recipient	1	0	9

[*]Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.17 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% of recipient's predicted mean	782	484	614
(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% of recipient's predicted mean	0	2	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	0	0	13
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,* age of donor ≥ age of recipient	0	0	2

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.18 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% of recipient's predicted mean	330	136	69
(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% of recipient's predicted mean	5	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	1	1	7
(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	2
(A) AFU of donor ≤ Age of recipient*	0	0	2

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.19 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% of recipient's predicted mean	217	216	167
(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% of recipient's predicted mean	1	0	9
(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	4
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,* age of donor ≥ age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient*	0	0	1

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.20 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% of recipient's predicted mean	123	105	62
(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% of recipient's predicted mean	3	1	6
(A) Age of donor = Age of recipient (B) Donor's predicted mean within 5% of recipient's predicted mean	0	0	0
(A) Age of donor = Age of recipient	4	0	2
(A) AFU of donor ≤ age of recipient (for overall hallucinogens),* age of donor ≥ age of recipient	0	0	1
(A) AFU of donor ≤ age of recipient (for overall hallucinogens)*	0	0	4

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.21 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% of recipient's predicted mean	399	484	614
(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% of recipient's predicted mean	11	2	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	6	0	13
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,* age of donor ≥ age of recipient	0	0	2

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.22 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% of recipient's predicted mean	55	75	48
(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% of recipient's predicted mean	5	1	19
(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	8	1	13
(A) Age of donor = Age of recipient	0	0	0
(A) AFU of donor ≤ age of recipient,* age of donor ≥ age of recipient	0	0	4

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.23 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% of recipient's predicted mean	19	18	16
(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% of recipient's predicted mean	4	5	10
(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	13	5	10
(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	3

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.24 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% of recipient's predicted mean	127	100	56
(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% of recipient's predicted mean	10	6	21
(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% of recipient's predicted mean	1	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)	12	5	9
(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)	1	0	4
(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)*	2	0	1

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.25 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% of recipient's predicted mean	37	68	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 (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% of recipient's predicted mean	2	4	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 (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% 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)	3	1	7
(A) Age of donor = Age of recipient (B) Donor agrees with recipient with respect to lifetime use for crack (checked only if recipient is a nonrespondent for crack AFU)	0	0	1
(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	2

*Although this is a logical constraint, it is included for the sake of clarity.

Exhibit I.26 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% of recipient's predicted mean	4	11	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 (C) Donor's predicted mean within 5% of recipient's predicted mean	0	2	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	3	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	2	0	0
(A) AFU of donor ≤ age of recipient*	0	0	1

*Although this is a logical constraint, it is included for the sake of clarity.

I.3 Health Insurance Variables

The imputation of the health insurance variables was also done separately for four age groups: 12 to17, 18 to 25, 26 to 64, and 65 or older. The two imputation-revised variables were created using a multivariate predictive mean neighborhood (MPMN) method. **Exhibit I.27** shows the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

Exhibit I.27 Health Insurance (IRINSUR) and Private Health Insurance (IRPINSUR)

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% of recipient's predicted means	849	289	112	18

I.4 Income Variables

The imputation of the income variables was also done separately for four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. It was done in two phases. The first phase, the "binary variable phase," involved the imputation of all the binary income variables, as well as the number of months on welfare. This phase was done using an MPMN method. The second phase, the "specific category phase," consisted of imputing more specific income categories for the respondent and the respondent's family in the household. This phase was done using a univariate predictive mean neighborhood (UPMN) method. **Exhibits I.28** and **I.29** show the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

Exhibit I.28 Income: Binary Variable Phase

	Frequency			
Likeness Constraints	**12-17**	**18-25**	**26-64**	**65+**
(A) Age of donor = Age of recipient (B) Donor's values for edited binary income variables are the same as recipient's values (when nonmissing)	1,502	809	310	127
(A) Age of donor is within 5 years of age of recipient (B) Donor's values for edited binary income variables are the same as recipient's values (when nonmissing)	47	34	172	14
(A) Age of donor is within 5 years of age of recipient (B) Same as (B) above, except that these checks are only done for those income sources for which the recipient is missing either the personal or the other family responses (C) If recipient's months on welfare response is missing, donor's values for the four welfare income source questions (PPMT, OFMPMT, PSVC, and OFMSVG) are the same as recipient's values (when nonmissing)	49	46	32	12
Use provisionally imputed values for whichever income variables are missing	0	2	0	2

Exhibit I.29 Income: Specific Category Phase

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
A) Donor's predicted mean within 10% of recipient's predicted mean B) PINC2 of donor = PINC2 of recipient, if nonmissing C) FINC2 of donor = FINC2 of recipient, if nonmissing	4,779	3,507	2,245	661
A) Donor's predicted mean within 10% of recipient's predicted mean B) FINC2 of donor ≥ PINC2 of recipient, if not missing* C) PINC2 of donor ≤ FINC2 of recipient, if not missing*	10	7	7	3
A) Donor's predicted mean within 20% of recipient's predicted mean B) FINC2 of donor ≥ PINC2 of recipient, if not missing* C) PINC2 of donor ≤ FINC2 of recipient, if not missing*	0	2	14	1
Take donor with predicted mean closest to recipent's predicted mean	1	1	3	0

*Although this is a logical constraint, it is included for the sake of clarity.

I.5 Household Roster-Derived Variables

The imputation of the roster variables was done separately for four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. Each of the four imputation-revised variables was created using a UPMN method. **Exhibits I.30** to **I.33** show the number of item nonrespondents who received values from donors meeting each set of likeness constraints.

Exhibit I.30 Total Number of Rostered People (IRHHSIZE)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
Donor's predicted mean within 10% of recipient's predicted mean	128	91	68	17

Exhibit I.31 Total Number of Kids under 18 (IRKID17)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 10% of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	247	207	159	27
(A) Donor's predicted mean within 20% of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	0	1	0	1
(A) Donor's predicted mean within 20% of recipient's predicted mean	0	2	1	0
Take donor with predicted mean closest to recipent's predicted mean	1	0	0	2

Exhibit I.32 Total Number of People 65 or Older (IRHH65)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 10% of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	488	298	167	21
(A) Donor's predicted mean within 20% of recipient's predicted mean (B) IRHHSIZE of donor = IRHHSIZE of recipient	1	0	0	0
(A) Donor's predicted mean within 20% of recipient's predicted mean	1	0	0	1
Take donor with predicted mean closest to recipent's predicted mean	0	0	0	1

Exhibit I.33 Indicator of Whether the Respondent Has Family Members in Household (IRFAMSKP)

Likeness Constraints	Frequency			
	12-17	18-25	26-64	65+
(A) Donor's predicted mean within 10% of recipient's predicted mean (B) IRKID17 of donor = IRKID17 of recipient	179	204	141	20
(A) Donor's predicted mean within 20% of recipient's predicted mean (B) IRKID17 of donor = IRKID17 of recipient	0	1	0	1
(A) Donor's predicted mean within 20% of recipient's predicted mean	1	1	0	1
Take donor with predicted mean closest to recipent's predicted mean	0	0	0	1

Appendix J

Missingness Patterns, Logical Constraints on MPMN Sets of Eligible Donors, and Portions of the Predictive Mean Vector Used in MPMNs

This page intentionally left blank.

Appendix J

Missingness Patterns, Logical Constraints on MPMN Sets of Eligible Donors, and Portions of the Predictive Mean Vector Used in MPMNs

J.1 Introduction

In the creation of the multivariate predictive mean neighborhoods (MPMNs) for recency of use, 12-month frequency of use, 30-day frequency of use, and 30-day binge drinking frequency, the set of donors was restricted prior to the calculation of the Mahalanobis distances. All models and imputations were conducted within the three age groups: 12 to 17 year olds, 18 to 25 year olds, and respondents 26 years of age or older. Donors for a given recipient were restricted according to logical and likeness constraints, where the likeness constraints were loosened if donors could not be found, but logical constraints could not be loosened. The number of respondents for whom donors were found under various likeness constraints is summarized in **Appendix I**. The logical constraints that an item respondent was required to satisfy to be a donor depended on the drug in question and the pattern of missingness.

In **Section J.2's** exhibits, the various missingness patterns are laid out, together with the frequency with which each pattern occurred. A few things to note on the exhibits are as follows.[1] In the missingness pattern section, no entry in the columns indicates all information is available; an entry of "Missing" indicates all information is missing. Other entries in the missingness pattern section give the information that is available, indicating that the information is partially missing. However, if the entry is in parentheses, all information is present and was thought to be useful for the reader. Please note that pain relievers, sedatives, and tranquilizers have identical missingness patterns and are therefore presented in the same table. The logical constraints associated with each missingness pattern that were used to restrict the set of donors are also given in **Section J.2's** exhibits.

The portion of the full predictive mean vector that was used in the calculation of the Mahalanobis distance (and the eventual determination of the hot-deck neighborhood) depended upon the missingness pattern that existed. **Section J.3's** exhibits give the various missingness patterns for each drug, the frequency which each pattern occurred, and the portions of the full predictive mean vector, with appropriate adjustments, that were used to create MPMNs for respondents with each respective missingness pattern.

---

[1] Many exhibits abbreviate certain words. "Recency" is an abbreviation for "Recency of Use," "Frequency" or "Freq" is an abbreviation for "Frequency of Use," "30-day binge drink" or "DR5DAY" is an abbreviation for the "number of days in the past 30 days when the respondent consumed of five or more alcoholic drinks."

This page intentionally left blank.

Section J.2

Exhibits Showing Missingness Patterns and the Restrictions on the Set of Potential Donors

This page intentionally left blank.

Exhibit J.1 Constraints for Tobacco (Cigarettes and Cigars)

Constraint #	Logical Constraint
Tob1	If the difference between the recipient's current age and his/her age at first use is 2 years or less, the recipient must have used within the past 3 years (a recency category of 1, 2, or 3)
Tob2	Recipient cannot be a past month user (recency cannot equal 1)
Tob3	Recipient must used drug within the past year (recency = 1 or 2)
Tob4	Recipient must be a past month user (recency = 1)
Tob5	If the recipient was never a daily user of cigarettes (CG15=2), the donor's 30-day cigarette frequency cannot equal 30
Tob6	If recipient's age at first use equals his/her current age, the donor's 30-day frequency (1) cannot be greater than the number of days between the recipient's interview date and his/her date of first drug use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his/her birthday (inclusive)

Exhibit J.2 Cigarette User Restrictions

Missingness Pattern			Number of Cases	Logical Constraints
#	Recency	30-Day Frequency		
1	Past year	Missing	14	(Tob1), (Tob5)
2	Missing (lifetime use imputed)	Missing	0	(Tob1), (Tob5)
2	Missing (lifetime use known)	Missing	154	
3	(Past month)	Missing	103	(Tob1), (Tob4), (Tob5), (Tob6)
4	Not past year		53	(Tob1), (Tob3), (Tob5)
5	Not past month		493	(Tob1), (Tob2), (Tob5)
6	30-day frequency logically assigned based on estimated value, no missing values.		710	(Tob1), (Tob5)
	Lifetime user, nothing missing		37,075	(None)
	Imputed to lifetime nonuse		0	(None)
	Lifetime nonuser, nothing missing		28,104	(None)

Exhibit J.3 Cigar User Restrictions

Missingness Pattern			Number of Cases	Logical Constraints
#	Recency	30-Day Frequency		
1	Past year	Missing	14	(Tob1)
2	Missing (Lifetime use imputed)	Missing	15	(Tob1)
2	Missing (Lifetime use known)	Missing	102	
3	(Past month)	Missing	33[1]	(Tob1), (Tob4), (Tob6)
4	Not past year		49	(Tob1), (Tob3)
5	Not past month		479	(Tob1), (Tob2)
6	30-day frequency logically assigned based on estimated value, no missing values.		150	(Tob1)
	Lifetime user, nothing missing		20,614	
	Imputed to lifetime nonuse		29	
	Lifetime nonuser, nothing missing		45221	

[1]For one case, the missingness pattern was changed from pattern #1 to pattern #3 because the interview and birth date indicated the respondent had to be a past month user.

Exhibit J.4 Constraints for Smokeless Tobacco (Chewing Tobacco and Snuff)

Constraint #	Description
SLT1	If the difference between the recipient's current age and his/her age at first chew use is 2 years or less, the recipient must have used chew within the past 3 years (a recency category of 1, 2, or 3)
SLT2	If the difference between the recipient's current age and his/her age at first snuff use is 2 years or less, the recipient must have used snuff within the past 3 years (a recency category of 1, 2, or 3)
SLT3	Donor's not a chew user, then recipient must also not be a chew user (and vice versa)
SLT4	Donor's not a snuff user, then recipient must also not be a snuff user (and vice versa)
SLT5	If recipient's age at first chew use equals his/her current age, the donor's 30-day chew frequency (1) cannot be greater than the number of days between the recipient's interview date and his/her date of first chew use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his/her birthday (inclusive)
SLT6	If recipient's age at first snuff use equals his/her current age, the donor's 30-day snuff frequency (1) cannot be greater than the number of days between the recipient's interview date and his/her date of first snuff use (inclusive) and (2) cannot be greater than the number of days between the recipient's interview date and his/her birthday (inclusive)
SLT7	Donor must be a past month chew user (chew recency = 1)
SLT8	Donor must be a past month snuff user (snuff recency = 1)
SLT9	Donor's snuff recency equal to recipient's snuff recency
SLT10	Donor's chew recency must equal recipient's chew recency
SLT11	Donor must have used chew within the past year (snuff recency = 1 or 2)
SLT12	Donor must have used snuff within the past year (chew recency = 1 or 2)
SLT13	Donor must be a past 3 years (but not past year) or lifetime (but not past 3 years) chew user (chew recency = 3 or 4)
SLT14	Donor must be a past 3 years (but not past year) or lifetime (but not past 3 years) snuff user (snuff recency = 3 or 4)
SLT15	Donor must be a past year (but not past month), past 3 years (but not past year) or lifetime (but not past 3 years) chew user (chew recency = 2, 3 or 4)
SLT16	Donor must be a past year (but not past month), past 3 years (but not past year) or lifetime (but not past 3 years) snuff user (snuff recency =2, 3 or 4)

Exhibit J.5 Smokeless Tobacco Users (Snuff and Chewing Tobacco) Restrictions

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints
1	(Past month)	(Past month)	Missing	Missing	32	(SLT1-SLT4), (SLT5-SLT8)
2	(Past month)		Missing		43	(SLT1-SLT4), (SLT5), (SLT7), (SLT9)
3		(Past month)		Missing[1]	36	(SLT1-SLT4), (SLT6), (SLT8), (SLT10)
4		Missing (Lifetime use known)		Missing	20	(SLT1-SLT4), (SLT6), (SLT10)
4		Missing (Lifetime use imputed)		Missing	0	(SLT1-SLT4), (SLT6), (SLT10)
5	(Past month)	Missing (Lifetime use known)	Missing	Missing	4	(SLT1-SLT4), (SLT5-SLT6), (SLT10)
5	(Past month)	Missing (Lifetime use imputed)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT10)
6	Missing (lifetime use known)		Missing		42	(SLT1-SLT4), (SLT5), (SLT9)
6	Missing (lifetime use imputed)		Missing		0	(SLT1-SLT4), (SLT5), (SLT9)
7	Missing (lifetime use known)	(Past month)	Missing	Missing	1	(SLT1-SLT4), (SLT5-SLT6), (SLT8)
7	Missing (lifetime use imputed)	(Past month)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6), (SLT8)
8		Past year		Missing	0	(SLT1-SLT4), (SLT10-SLT11)
9	Past year		Missing		7	(SLT1-SLT4), (SLT5), (SLT8), (SLT12)
10	Missing (lifetime use known)	Missing (Lifetime use known)	Missing	Missing	10	(SLT1-SLT4), (SLT5-SLT6)

Exhibit J.5 (continued)

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints
10	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6)
10	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6)
10	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	(SLT1-SLT4), (SLT5-SLT6)
11	Not past year				22	(SLT1-SLT4), (SLT8), (SLT13)
12		Not past year			12	(SLT1-SLT4), (SLT10), (SLT14)
13	Not past year	Not past year			2	(SLT1-SLT4), (SLT13-SLT14)
14	Not past month				192	(SLT1-SLT4), (SLT9), (SLT15)
15		Not past month			87	(SLT1-SLT4), (SLT10), (SLT16)
16	Not past month	Not past month			20	(SLT1-SLT4), (SLT15-SLT16)
17	Not past month	(Past month)		Missing	0	(SLT1-SLT4), (SLT6), (SLT8), (SLT15)
18	(Past month)	Not past month	Missing		1	(SLT1-SLT4), (SLT5), (SLT7), (SLT16)
19	Not past month	Missing (lifetime use known)		Missing	2	(SLT1-SLT4), (SLT6), (SLT15)
19	Not past month	Missing (lifetime use imputed)		Missing	0	(SLT1-SLT4), (SLT6), (SLT15)
20	Missing (lifetime use known)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT16)
20	Missing (lifetime use imputed)	Not past month	Missing		0	(SLT1-SLT4), (SLT5), (SLT16)
21	Not past month	Not past year			0	(SLT1-SLT4), (SLT14-SLT15)
22	Not past year	Not past month			0	(SLT1-SLT4), (SLT13), (SLT16)

Exhibit J.5 (continued)

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Logical Constraints
23	(Lifetime use of snuff, chewing tobacco, or both missing in raw data. Missing values imputed to nonuse in lifetime imputation; nothing missing at this point in sequence)				0	(SLT1-SLT4)
	Lifetime user, nothing missing				12,256	
	Imputed to lifetime nonuse				59	
	Lifetime nonuser, nothing missing				53,858	

[1]For one case, the missingness pattern was changed from pattern #8 to pattern #3 because the interview and birth date indicated the respondent had to be a past month user.

Exhibit J.6 Pipe User Restrictions

Missingness Pattern		Number of Cases	Constraints
#	Recency		
1	Missing (lifetime use imputed)	7	(None)
1	Missing (lifetime use known)	13	(None)
	Lifetime user, nothing missing	6,351	
	Imputed to lifetime nonuse	30	
	Lifetime nonuser, nothing missing	60,305	

Exhibit J.7 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)

Exhibit J.7 (continued)

Drug	Constraint #	Constraint
Alc, Mrj, Inh	C11	If recipient's age at first use equals his/her current age, the donor's 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first drug use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Alc, Mrj, Inh	C12	If recipient's age at first use equals his/her current age, (1) recipient's donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Alc, Mrj, Inh	C13	Recipient's estimated 30-day frequency is not given/legitimately skipped (estimated frequency not equal to 1-6)
Alc, Mrj, Inh	C14	If recipient's age at first use equals his/her current age, (1) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (-29) and (2) donor's proportion of past year use * recipient's max number of days could have used in past year cannot be greater than the recipient's days between the interview date and birthday (-29)
Alc, Mrj, Inh, Anl, Trn. Sed	C15	Donor must be a past year (but not past month) user (recency = 2)
Alc, Mrj, Inh	C16	Donor's DR5DAY values is less than recipient's 30-day frequency
Alc, Mrj, Inh	C17	If recipient's age at first use equals his/her current age, (1) donor's DR5DAY must be less than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's DR5DAY must be less than recipient's days between his/her interview date and birthday (+1)
Alc, Mrj, Inh, Anl, Trn. Sed	C18	Donor must be a past month or past year (but not past month) use (recency = 1 or 2)
Alc, Mrj, Inh	C19	Donor's proportion of past year use * recipient's max number of days could have used in past year greater than donor's 30-day frequency
Alc, Mrj, Inh, Her	C20	If recipient's age at first use equals his/her current age, (1) donor's proportion of past year used * recipient's max number of days could have used in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (-365) and (2) donor's proportion of past year used * recipient's max number of days could have used in past year cannot be greater than the recipient's days between his/her interview date and birthday (-365)
Alc, Mrj, Inh, Her	C21	Donor's proportion of past year used * recipient's max number of days could have used in past year cannot be greater than recipient's max number of days could have used in past year (30 + 30-day frequency)

Exhibit J.8 Alcohol User Restrictions

Missingness Pattern						
#	**Recency**	**12-Month Freq.**	**30-Day Freq.**	**30-Day Binge Drink**	**Number of Cases**	**Logical Constraints**
1	(Past month)	Missing	Missing		90	(C1-C13)
2	(Past month)		Missing		402	(C5-C8), (C10), (C11), C13
3	(Past month)	Missing			232	(C1-C4), (C10), (C12)
4	(Past year but not past month)	Missing			333	(C1-C3), (C14), (C15)
5	(Past month)			Missing	947	(C10), (C16), (C17)
6	(Past month)		Missing	Missing[1]	204	(C5-C7), (C10), (C11), (C13)
7	(Past month)	Missing		Missing	45	(C1-C4), (C10), (C12), (C16), (C17)
8	(Past month)	Missing	Missing	Missing	147	(C1-C4), (C5-C7), (C9-C13)
9	Past Year		Missing	Missing	454	(C5-C7), (C11), (C13, C15)
10	Past year	Missing	Missing	Missing	52	(C1-C3), (C5-C9), (C11-C14), (C18)
11	Lifetime (imputed)	Missing	Missing	Missing	0	(C1-C7), (C9), (C11-C14) (C1-C70, (C9), (C11-C14)
11	Lifetime (known)	Missing	Missing	Missing	904	
	(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.)					
	Lifetime user, nothing missing				42,413	
	Imputed to lifetime nonuse				21	
	Lifetime nonuser, nothing missing				20,462	

[1]For two cases, the missingness pattern was changed from pattern #9 to pattern #6. The first case changed because the interview and birth date indicated the respondent had to be a past month user. The second case changed because the month of first use and the 12-month frequency indicated the respondent had to be a past month user.

Exhibit J.9 Marijuana User Restrictions

Missingness Pattern					
#	**Recency**	**12-Month Frequency**	**30-Day Frequency**	**Number of Cases**	**Constraints**
1	(Past month)	Missing	Missing	60	(C1-C7), (C9-C13)
2	(Past month)		Missing	54[1]	(C5-C7), (C10), (C11), (C13)
3	(Past month)	Missing		54	(C1-C4), (C10), (C12)
4	(Past year but not past month)	Missing		103	(C1-C3), (C13), (C14)
5	Past year		Missing	112	(C5-C7), (C11), (C13), (C18)
6	Past year	Missing	Missing	80	(C1-C3), (C5-C7), (C9), (C11-C14), (C18), (C19)
7	Missing (lifetime use imputed)	Missing	Missing	0	(C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20) (C1-C3), (C5-C7), (C9), (C11-C14), (C19),(C20)
7	Missing (lifetime use known)	Missing	Missing	585	
	Lifetime user, nothing missing			21,613	
	Imputed to lifetime nonuse			135	
	Lifetime nonuser, nothing missing			43,910	

[1]For one case, the missingness pattern was changed from pattern #5 to pattern #2 because the month of first use and the 12-month frequency indicated the respondent had to be a past month user.

Exhibit J.10 Inhalants User Restrictions

Missingness Pattern				Number of Cases	Constraints
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	10	(C1-C7), (C10), (13)
2	(Past month)		Missing	12	(C6-C8), (C10), (C13)
3	(Past month)	Missing		6	(C1-C4), (C10)
4	(Past year not past month)	Missing		36	(C1-C3), (C18)
5	Past year		Missing	35	(C5-C7), (C9),(C13), (C18)
6	Past year	Missing	Missing	3	(C1-C3), (C5-C7), (C9), (C13), (C18)
7	Missing (lifetime use imputed)	Missing	Missing	0	(C1-C3), (C5-C7), (C9), (C13)
7	Missing (lifetime use known)	Missing	Missing	436	(C1-C3), (C5-C7), (C9), (C13)
	Lifetime user, nothing missing			6,412	
	Imputed to lifetime nonuse			172	
	Lifetime nonuser, nothing missing			59,584	

Exhibit J.11 Heroin User Restrictions

Missingness Pattern				Number of Cases	Constraints
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	0	(C1-C7), (C9), (C10-C13), (C21)
2	(Past month)		Missing	1	(C5-C7), (C10), (C13)
3	(Past month)	Missing		0	(C1-C4), (C10), (C21)
4	(Past year but not past month)	Missing		2	(C1-C3), (C15)
5	Past year		Missing	2	(C5-C7), (C9), (C13), (C18)
6	Past year	Missing	Missing	4	(C1-C3), (C5-C7), (C9), (C13), (C18), (C21)
7	Missing (lifetime use imputed)	Missing	Missing	0	(C1-C3), (C5-C7), (C9), (C13), (C21)
7	Missing (lifetime use known)	Missing	Missing	31	(C1-C3), (C5-C7), (C9), (C13), (C21)
	Lifetime user, nothing missing			685	
	Imputed to lifetime nonuse			74	
	Lifetime nonuser, nothing missing			65,907	

Exhibit J.12 Users of Pain Relievers, Tranquilizers, and Sedatives

Missingness Pattern				
#	Recency	12-Month Frequency	Number of Cases	Constraints
1	(Past month)	Missing	Pain relievers: 77	(C1-C3), (C10)
			Tranquilizers: 11	
			Sedatives: 4	
2	(Past year but not past month)	Missing	Pain relievers: 89	(C1-C3), (C15)
			Tranquilizers: 21	
			Sedatives: 5	
3	Past year		Pain relievers: 4	(C18)
			Tranquilizers: 0	
			Sedatives: 0	
4	Past year	Missing	Pain relievers: 9	(C1-C3), (C18)
			Tranquilizers: 1	
			Sedatives: 0	
5	Missing (lifetime use imputed)	Missing	Pain relievers: 0	(C1-C3), (C18) (C1-C3), (C18)
			Tranquilizers: 0	
			Sedatives: 0	
5	Missing (lifetime use known)	Missing	Pain relievers: 597	
			Tranquilizers: 214	
			Sedatives: 87	
	Lifetime user, nothing missing		Pain relievers: 6,313[1]	
			Tranquilizers: 3,341	
			Sedatives: 1,275	
	Imputed to lifetime nonuse		Pain relievers: 409	
			Tranquilizers: 276	
			Sedatives: 300	
	Lifetime nonuser, nothing missing		Pain relievers: 59,208	
			Tranquilizers: 62,842	
			Sedatives: 65,035	

[1]For one case, the missingness pattern was changed from pattern #3 to "lifetime user, nothing missing" because the interview date, birth date, and 12-month frequency indicated the respondent had to be a past month user and the 12-month frequency was not missing.

Exhibit J.13 Constraints for Cocaine and Crack

Constraint #	Constraint
Coc1	Donor must be a past month cocaine user (cocaine recency = 1)
Coc2	Donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year must be less than (or equal) the recipient's maximum possible past year cocaine frequency of use. The recipient's maximum possible cocaine frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used cocaine, as determined by the month of first use (2) if the maximum period the recipient could have used cocaine is greater than 30, but the recipient is a past month cocaine user with a nonmissing 30-day frequency, the past year cocaine frequency must be less than or equal to the maximum period (the number of days the recipient did not use in the past month) (3) if the recipient is not a past cocaine month user, the past year cocaine frequency must be less than or equal to the maximum period (30)
Coc3	Donor's proportion of past year cocaine use * recipient's min number of days could have used cocaine in past year must be greater than (or equal) the recipient's minimum possible past year cocaine frequency of use. The recipient's minimum possible cocaine frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month cocaine user, it must be at least as much as the 30-day freq (2) if the recipient is not a past month cocaine user but a past year cocaine user, it must be at least 1
Coc4	(Recipient's proportion of past year cocaine use * max number of days could have used cocaine in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Coc5	(Donor's proportion of past year cocaine use * recipient's number of days could have used cocaine in past year) greater than or equal to 30-day use
Coc6	Donor's 30-day cocaine use less than number of days between recipient's interview date and birthday (+1)
Coc7	Donor's 30-day cocaine use less than the recipient's maximum number of days could have used in past 30 days
Coc8	Donor's 30-day cocaine use greater than the recipient's minimum number of days could have used in past 30 days
Coc9	If recipient's age at first cocaine use equals his/her current age, the donor's cocaine 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first cocaine use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)

Exhibit J.13 (continued)

Constraint #	Constraint
Coc10	If recipient's age at first cocaine use equals his/her current age, (1) recipient's donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's proportion of past year cocaine use* recipient's max number of days could have used cocaine in past year cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Coc11	Recipient's estimated cocaine 30-day frequency is not given/legitimately skipped (estimated cocaine frequency not equal to 1-6)
Coc12	Donor's crack recency equals recipient's crack recency
Coc13	Donor must be a past year (but not past month) cocaine user (cocaine recency = 2)
Coc14	If recipient's age at first cocaine use equals his/her current age, donor's proportion of past year cocaine use * recipient's max number of days could have used cocaine in past year cannot be greater than recipient's days between his/her interview date and date of first cocaine use (-29)
Coc15	Donor must be a past month or past year (but not past month) cocaine user (cocaine recency = 1 or 2)
Coc16	Donor must be a past month, past year (but not past month), or a lifetime (but not past year) cocaine user (cocaine recency = 1, 2, or 3**)**
Coc17	If recipient's age at first cocaine use equals his/her current age, donor cannot be a lifetime (but not past year) cocaine user (cocaine recency cannot equal 3**)**
Coc18	Donor's proportion of past year crack use * recipient's max number of days could have used crack in past year must be less than (or equal) the recipient's maximum possible past year crack frequency of use. The recipient's maximum possible crack frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used crack, as determined by the month of first use (2) if the maximum period the recipient could have used crack is greater than 30, but the recipient is a past month crack user with a nonmissing 30-day frequency, the past year crack frequency must be less than or equal to the maximum period (the number of days the recipient did not use in the past month) (3) if the recipient is not a past crack month user, the past year crack frequency must be less than or equal to the maximum period (30)
Coc19	Donor's proportion of past year crack use * recipient's min number of days could have used crack in past year must be greater than (or equal) the recipient's minimum possible past year crack frequency of use. The recipient's minimum possible crack frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month crack user, it must be at least as much as the 30-day freq (2) if the recipient is not a past month crack user but a past year crack user, it must be at least 1

Exhibit J.13 (continued)

Constraint #	Constraint
Coc20	(Recipient's proportion of past year crack use * max number of days could have used crack in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Coc21	(Donor's proportion of past year crack use * recipient's number of days could have used crack in past year) greater than or equal to 30-day use
Coc22	Donor's 30-day crack use less than number of days between recipient's interview date and birthday (+1)
Coc23	Donor's 30-day crack use less than the recipient's maximum number of days could have used in past 30 days
Coc24	Donor's 30-day crack use greater than the recipient's minimum number of days could have used in past 30 days
Coc25	If recipient's age at first crack use equals his/her current age, the donor's crack 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first crack use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1**)**
Coc26	If recipient's age at first crack use equals his/her current age, (1) recipient's donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Coc27	Recipient's estimated 30-day crack frequency is not given/legitimately skipped (estimated crack frequency not equal to 1-6)
Coc28	Donor must be a past month crack user (crack recency = 1)
Coc29	Donor must be a past month or past year (not past month) crack user (crack recency = 1, 2)
Coc30	Donor must be a past month, past year (not past month), or lifetime (but not past year) crack user (crack recency = 1, 2)
Coc31	Donor's cocaine recency must equal recipient's cocaine recency or donor's cocaine recency must equal recipient's cocaine recency (10)
Coc32	If recipient's age at first crack use equals his/her current age donor cannot be a lifetime (but not past year) crack user (crack recency cannot equal 3**)**
Coc33	Donor must be a past year (but not past month) crack user (crack recency = 2)
Coc34	If recipient's age at first crack use equals his/her current age, donor's proportion of past year crack use * recipient's max number of days could have used crack in past year cannot be greater than recipient's days between his/her interview date and date of first crack use (-29)

Exhibit J.14 Cocaine User Restrictions

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
1	(Past month)		Missing		Missing		7	(Coc1-Coc12)
2	(Past month)				Missing		22	(Coc1), (Coc6-Coc9), (Coc11-Coc12)
3	(Past month)		Missing				4	(Coc2-Coc4), (Coc10), (Coc12)
4	(Past year not past month)		Missing				21	(Coc2-Coc4), (Coc12-Coc14)
5	Past year				Missing		22	(Coc6-Coc9), (Coc11-Coc12), (Coc15)
6	Past year		Missing		Missing		30	(Coc2-Coc12), (Coc15)
7	Missing (lifetime use known)		Missing		Missing		131	(Coc2-Coc12), (Coc16-Coc17)
7	Missing (lifetime use imputed)		Missing		Missing		13	
8	(Past month)	(Past month)		Missing		Missing	1	(Coc1), (Coc18-Coc27)
9	(Past month)	(Past month)					4	(Coc1), (Coc22-Coc25), (Coc27-Coc28)

Exhibit J.14 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
10	(Past year not missing)	Past year (not missing)		Missing			3	(Coc1), (Coc18-Coc20), (Coc26), (Coc28)
12	(Past month)	Past year				Missing	3	(Coc1), (Coc22-Coc25), (Coc27), (Coc29)
13	(Past month)	Past year		Missing		Missing	1	(Coc1), (Coc18-Coc27), (Coc29)
14	(Past month)	Missing (Lifetime use known)		Missing		Missing	4	(Coc16), (Coc18-Coc26), (Coc30-Coc32)
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)
16	(Past month)	(Past year but not past month)	Missing	Missing			1	(Coc1-Coc4), (Coc10), (Coc18-Coc20), (Coc26), (Coc33)

Exhibit J.14 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			2	(Coc2-Coc4), (Coc14), (Coc18-Coc20), (Coc33-Coc34)
18	(Past month)	(Past month)			Missing	Missing	1	(Coc1), (Coc6-Coc9), (Coc11), (Coc22-Coc25), (Coc27-Coc28)
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	1	(Coc1-Coc11), (Coc18-Coc28)
20	(Past month)	(Past month)	Missing		Missing	Missing	0	(Coc1-Coc11), (Coc16), (Coc22-Coc25), (Coc27-Coc28)
21	(Past month)	(Past month)		Missing	Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc28)
22	(Past month)	(Past month)	Missing	Missing	Missing		1	(Coc1-Coc11), (Coc18-Coc21), (Coc26), (Coc28)

Exhibit J.14 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
23	(Past month)	(Past month not past year)	Missing	Missing	Missing		1	(Coc1-Coc11), (Coc18-Coc20), (Coc33), (Coc34)
24	(Past month)	(Past month)	Missing	Missing		Missing	0	(Coc1-Coc4), (Coc10), (Coc18-Coc26), (Coc28)
25	(Past month)	(Past month)		Missing	Missing		0	(Coc1), (Coc6-Coc9), (Coc18-Coc20), (Coc26), (Coc28)
26	(Past month)	(Past year not past month)		Missing	Missing		0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc 20), (Coc26), (Coc33)
27	(Past month)	(Past month)	Missing			Missing	0	(Coc1-Coc4), (Coc10), (Coc22-Coc25), (Coc27-Coc28)
28	Past year	Past year			Missing	Missing	4	(Coc6-Coc9), (Coc11), (Coc15), (Coc22-Coc25), (Coc27), (Coc29)

Exhibit J.14 (continued)

Missingness Pattern							Number of Cases	Constraints
#	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.		
29	Past year	Past year	Missing		Missing	Missing	0	(Coc3-Coc11), (Coc15), (Coc21-Coc25), (Coc27), (Coc29)
30	Past year	Past year		Missing	Missing	Missing	4	(Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc29)
31	Past year	Past year	Missing	Missing	Missing	Missing	5	(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc29)
32	Past year	Missing (lifetime use known)		Missing	Missing	Missing	12	(Coc1), (Coc6-Coc9), (Coc11), (Coc15), (Coc18-Coc27), (Coc30)
32	Past year	Missing (lifetime use imputed)		Missing	Missing	Missing	0	
33	Past year	Missing (lifetime use known)	Missing	Missing	Missing	Missing	2	(Coc2-Coc11), (Coc15), (Coc18-Coc27), (Coc30), (Coc32)
33	Past year	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0	

Exhibit J.14 (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
34	(Past month)	Missing (lifetime use known)		Missing	Missing	Missing	0	(Coc1), (Coc6-Coc9), (Coc11), (Coc18-Coc27), (Coc30), (Coc32)
34	(Past month)	Missing (lifetime use imputed)		Missing	Missing	Missing	0	
35	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	(Coc1-Coc11), (Coc18-Coc27), (Coc30)
35	(Past month)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0	
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	27	(Coc2-Coc11), (Coc16-Coc27), (Coc30)
36	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	10	
36	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	Missing	Missing	0	
	Cocaine was logically assigned one of the "past year" recency categories based on the response to the crack question. There were no missing values.							
	Lifetime user, nothing missing						5,564	
	Imputed to lifetime nonuse						91	
	Lifetime nonuser, nothing missing						60,714	

Exhibit J.15 Constraints for Hallucinogens (Including LSD and PCP)

Con-straint #	Constraint
Hal1	Donor must be a LSD user (LSD recency not equal to 91)
Hal2	Donor's hallucinogen recency must equal recipient's hallucinogen recency or donor's hallucinogen recency must equal recipient's hallucinogen recency (10)
Hal3	Donor's PCP recency must equal recipient's PCP recency
Hal4	Donor must be a PCP user (PCP recency not equal to 91)
Hal5	Donor's LSD recency must equal recipient's LSD recency
Hal6	Donor must be a LSD and PCP user (LSD and PCP recencies not equal to 91)
Hal7	Donor's must be a past month hallucinogens user (hallucinogen recency = 1)
Hal8	Donor's proportion of past year hallucinogen use * recipient's max number of days could have used hallucinogens in past year must be less than (or equal) the recipient's maximum possible past year hallucinogen frequency of use. The recipient's maximum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used hallucinogens, as determined by the month of first use (2) if the maximum period the recipient could have used hallucinogens is greater than 30, but the recipient is a past month user with a nonmissing 30-day hallucinogen frequency, the past year hallucinogen frequency must be less than or equal to the maximum period (the number of days the recipient did not use hallucinogens in the past month) (3) if the recipient is not a past month hallucinogen user, the past year hallucinogen frequency must be less than or equal to the maximum period (30)
Hal9	Donor's proportion of past year hallucinogen use * recipient's min number of days could have used hallucinogens in past year must be greater than (or equal) the recipient's minimum possible past year hallucinogen frequency of use. The recipient's minimum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month hallucinogen user, it must be at least as much as the hallucinogen 30-day freq (2) if the recipient is not a past month hallucinogen user but a past year hallucinogen user, it must be at least 1
Hal10	(Recipient's proportion of past year hallucinogen use * max number of days could have used hallucinogens in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Hal11	Donor's 30-day hallucinogen use less than number of days between recipient's interview date and birthday (+1)
Ha12	Donor's 30-day hallucinogen use less than the recipient's maximum number of days could have used hallucinogens in past 30 days

Exhibit J.15 (continued)

Con-straint #	Constraint
Hal13	Donor's 30-day hallucinogen use greater than the recipient's minimum number of days could have used hallucinogens in past 30 days
Hal14	Donor must be a hallucinogen past year (but not past month) or past month user (hallucinogen recency = 1 or 2)
Hal15	Donor must be a LSD past year (but not past month) or past month user (LSD recency = 1 or 2)
Hal16	Donor must be a PCP past year (but not past month) or past month user (PCP recency = 1 or 2)
Hal17	Donor must be a LSD user (LSD recency = 1, 2, or 3)
Hal18	Donor must be a PCP user (PCP recency = 1, 2, or 3)
Hal19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
Hal20	Donor must not be a LSD past year (but not past month) or past month user (LSD recency not equal to 1 or 2)
Hal21	Donor must not be a PCP past year (but not past month) or past month user (PCP recency not equal to 1 or 2)

Exhibit J.16 Hallucinogen User Restrictions (Including LSD and PCP)

Missingness Pattern						Number of Cases	Constraints
#	Hallucinogen Recency	LSD Recency	PCP Recency	Hallucinogen 12-Mo. Freq.	Hallucinogen 30-Day Freq.		
1		Missing (lifetime use known)				2	(Hal1-Hal3)
1		Missing (lifetime use imputed)				1	
2			Missing (lifetime use known)			2	(Hal3), (Hal4-Hal5)
2			Missing (lifetime use imputed)			2	
3		Missing (lifetime use known)	Missing (lifetime use known)			0	(Hal3), (Hal6)
3		Missing (lifetime use imputed)	Missing (lifetime use imputed)			0	
3		Missing (lifetime use known)	Missing (lifetime use imputed)			0	
3		Missing (lifetime use imputed)	Missing (lifetime use known)			0	
4	(Past month)			Missing	Missing	13	(Hal7-Hal13)
5	(Past month)				Missing	18	(Hal7), (Hal11-Hal13)
6	(Past year)			Missing		91	(Hal2-Hal3), (Hal5), (Hal8-Hal10), (Hal14)

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
7	(Past month)	Missing (lifetime use known)			Missing	0	(Hal1), (Hal3), (Hal7), (Hal11-Hal13)
7	(Past month)	Missing (lifetime use imputed)			Missing	0	
8	(Past month)		Missing (lifetime use known)		Missing	0	(Hal4-Hal5), (Hal7), (Hal11-Hal13)
8	(Past month)		Missing (lifetime use imputed)		Missing	0	
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	1	(Hal6-Hal7), (Hal11-Hal13)
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0	(Hal6-Hal7), (Hal11-Hal13)
9	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0	
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0	
10	(Past month or Past month not past year)	Missing (lifetime use known)		Missing		0	(Hal1-Hal3), (Hal8-Hal10), (Hal14)

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	**Hallu-cinogen Recency**	**LSD Recency**	**PCP Recency**	**Hallu-cinogen 12-Mo. Freq.**	**Hallu-cinogen 30-Day Freq.**		
10	(Past month or Past month not past year)	Missing (lifetime use imputed)		Missing		0	
11	(Past month or Past month not past year)		Missing (lifetime use known)	Missing		0	(Hal2), (Hal4-Hal5), (Hal8-Hal10), (Hal14)
11	(Past month or Past month not past year)		Missing (lifetime use imputed)	Missing		0	
12	(Past month or Past month not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	(Hal2), (Hal6), (Hal8-Hal10), (Hal14)
12	Past year (not missing)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing		0	
12	Past year (not missing)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing		0	
12	(Past month or Past month not past year)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing		0	
13	(Past month)	Missing (lifetime use known)		Missing	Missing	0	(Hal1), (Hal3), (Hal7-13)

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
13	(Past month)	Missing (lifetime use imputed)		Missing	Missing	0	
14	(Past month)		Missing (lifetime use known)	Missing	Missing	1	(Hal4-Hal5), (Hal7-Hal13)
14	(Past month)		Missing (lifetime use imputed)	Missing	Missing	0	
15	(Past month	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal6-Hal13)
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
15	(Past month)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	
16	Past year	(Not past month)	(Not past month)		Missing	18	(Hal3), (Hal5), (Hal11-Hal14)
17	Past year	(Not past month)	(Not past month)	Missing	Missing	6	(Hal3), (Hal5), (Hal8-Hal14)
18	Past year	Past year	(Not past month)		Missing	7	(Hal3), (Hal11-Hal15)
19	Past year	(Not past month)	Past year		Missing	1	(Hal5), (Hal11-Hal14), (Hal16)
20	Past year	Past year	Past year		Missing	0	(Hal11-Hall16)

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
21	Past year	Missing (lifetime use known)	(Not past month)		Missing	6	(Hal3), (Hal11-Hal14), (Hal17)
21	Past year	Missing (lifetime use imputed)	(Not past month)		Missing	0	
22	Past year	(Not past month)	Missing (lifetime use known)		Missing	5	(Hal5), (Hal11-Hal14), (Hall18)
22	Past year	(Not past month)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	(Hal8-Hal14), (Hal17-Hal18)
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)		Missing	0	
23	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)		Missing	0	
24	Past year	Past year	(Not past month)	Missing	Missing	0	(Hal3), (Hal8-Hal15)
25	Past year	(Not past month)	Past year	Missing	Missing	0	(Hal5), (Hal8-Hal14), (Hal16)

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
26	Past year	Past year	Past year	Missing	Missing	0	(Hal8-Hal16)
27	Past year	Missing (lifetime use known)	(Not past month)	Missing	Missing	3	(Hal3), (Hal8-Hal14), (Hal17)
27	Past year	Missing (lifetime use imputed)	(Not past month)	Missing	Missing	0	
28	Past year	(Not past month)	Missing (lifetime use known)	Missing	Missing	2	(Hal5), (Hal8-Hal14), (Hal18)
28	Past year	(Not past month)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	(Hal8-Hal14), (Hal17-Hal18)
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	
29	Past year	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	
30	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	151	(Hal8-Hal13), (Hal19-Hal21)
30	Missing (lifetime use imputed)	(Not past year)	(Not past year)	Missing	Missing	20	

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
31	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	114	(Hal8-Hal13), (Hal17), (Hal19), (Hal21)
31	Missing (lifetime use imputed)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	13	
31	Missing (lifetime use known)	Missing (lifetime use imputed)	(Not past year)	Missing	Missing	2	
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	13	(Hal8-Hal13), (Hal18-Hal20)
32	Missing (lifetime use imputed)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	3	
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use imputed)	Missing	Missing	0	
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	20	(Hal8-Hal13), (Hal17-Hal19)
33	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	2	
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	Missing	0	
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	Missing	0	

Exhibit J.16 (continued)

Missingness Pattern						Number of Cases	Constraints
#	Hallu-cinogen Recency	LSD Recency	PCP Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.		
33	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing (lifetime use known)	Missing	Missing	0	
	Hallucinogens was logically assigned one of the "past year" recency categories based on the response to the LSD and/or PCP recency. There were no missing values.						
	Lifetime user, nothing missing					7,476	
	Imputed to lifetime nonuse					386	
	Lifetime nonuser, nothing missing					58,327	

Exhibit J.17 Constraints for Stimulants and Methamphetamines

Constraint #	Constraint
Stm1	Donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year must be less than (or equal) the recipient's maximum possible past year stimulants frequency of use. The recipient's maximum possible stimulants frequency of use in the past year is limited by the following factors: (1) it must be less or equal to than the maximum period the recipient could have used stimulants, as determined by the month of first use (2) if the maximum period the recipient could have used stimulants is greater than 30, but the recipient is a past month stimulants user with a nonmissing 30-day frequency, the past year stimulants frequency must be less than or equal to the maximum period (the number of days the recipient did not use in the past month) (3) if the recipient is not a past stimulants month user, the past year stimulants frequency must be less than or equal to the maximum period (30)
Stm2	Donor's proportion of past year stimulants use * recipient's min number of days could have used stimulants in past year must be greater than (or equal) the recipient's minimum possible past year stimulants frequency of use. The recipient's minimum possible stimulants frequency of use in the past year is limited by the following factors: (1) if the recipient is a past month stimulants user, it must be at least as much as the 30-day freq (2) if the recipient is not a past month stimulants user but a past year stimulants user, it must be at least 1.
Stm3	(Recipient's proportion of past year stimulants use * max number of days could have used stimulants in past year) less than or equal to the number of days between recipient's interview date and birthday (+1)
Stm4	Donor must be a past month stimulant user (stimulant recency = 1)
Stm5	Donor's meth recency equals the recipient's meth recency
Stm6	If recipient's age at first stimulants use equals his/her current age, (1) recipient's donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (+1) and (2) donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Stm7	Donor must be a past year (but not past month) stimulant user (stimulant recency = 2)
Stm8	If recipient's age at first stimulants use equals his/her current age, (1) recipient's donor's proportion of past year stimulants use* recipient's max number of days could have used stimulants in past year cannot be greater than recipient's days between his/her interview date and date of first drug use (-29) and (2) donor's proportion of past year stimulants use * recipient's max number of days could have used stimulants in past year cannot be greater than the recipient's days between his/her interview date and birthday (-29)

Exhibit J.17 (continued)

Constraint #	Constraint
Stm9	Donor must be a past month or past year (but not past month) stimulant user (stimulants recency = 1 or 2)
Stm10	If recipient's age at first stimulants use equals his/her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first stimulants use (+1) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (+1)
Stm11	Donor's stimulants recency must equal recipient's stimulants recency or donor's stimulants recency must equal recipient's stimulants recency (10).
Stm12	Donor must be a past month, past year (but not past month), or lifetime (but not past year) meth user (meth recency = 1, 2, or 3)
Stm13	If the number of days between the recipient's interview and birthday (+1) is between 0 and 30, meth recency must not equal 2 or 3
Stm14	If the number of days between the recipient's interview and birthday (+1) is between 0 and 365, meth recency must not equal 3
Stm15	If recipient's age at first stimulants use equals his/her current age or the recipient's age at first meth use equals his/her current age or the recipient's number of days between his/her interview date and date at first meth use less than 30, the donor's recency must not equal 3
Stm16	If recipient's age at first stimulants use equals his/her current age, the donor's stimulants 30-day frequency (1) cannot be greater than the recipient's days between his/her interview date and date of first stimulants use (-29) and (2) cannot be greater than the recipient's days between his/her interview date and birthday (-29)
Stm17	Donor must be a past month or past year (but not past month) meth user (meth recency = 1 or 2)

Exhibit J.18 Stimulants User Restrictions (Includes Methamphetamines)

Missingness Pattern				Number of Cases	Constraints
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency		
1	(Past month)		Missing	100	(Stm1-Stm6)
2	(Past year but not past month)		Missing	215	(Stm1-Stm3), (Stm5), (Stm7-Stm8)
3	Past year			1	(Stm5), (Stm8-Stm10)
4	Past year		Missing	11	(Stm1-Stm3), (Stm5-Stm6), (Stm8-Stm9)
5	Missing (lifetime use known)		Missing	165	(Stm1-Stm3), (Stm5-Stm6), (Stm8)
5	Missing (lifetime use imputed)		Missing	15	
6		Missing (lifetime use known)		12	(Stm11-Stm15)
6		Missing (lifetime use imputed)		11	
7	(Past month)	Missing (lifetime use known)	Missing	1	(Stm1-Stm4), (Stm6), (Stm12-Stm15)
7	(Past month)	Missing (lifetime use imputed)	Missing	0	
8	(Past year not past month)	Missing (lifetime use known)	Missing	0	(Stm1-Stm3), (Stm7-Stm8), (Stm12-Stm15)
8	(Past year not past month)	Missing (lifetime use imputed)	Missing	0	
9	Past year	Missing (lifetime use known)		9	(Stm9-Stm10), (Stm12-16)
9	Past year	Missing (lifetime use imputed)		0	
10	Past year	Missing (lifetime use known)	Missing	0	(Stm1-Stm3), (Stm6), (Stm8-Stm9), (Stm12-Stm15)
10	Past year	Missing (lifetime use imputed)	Missing	0	
11	Past year (not missing)	Past year		1[1]	(Stm11), (Stm13), (Stm17)

Exhibit J.18 (continued)

Missingness Pattern				Number of Cases	Constraints
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency		
12	Past month	Past year	Missing	0	(Stm1-Stm4), (Stm6), (Stm9), (Stm13)
13	(Past year not past month)	Past year	Missing	0	(Stm1-Stm3), (Stm7-Stm8), (Stm13), (Stm17)
14	Past year	Past year		1[2]	(Stm9-Stm10), (Stm13), (Stm16-Stm17)
15	Past year	Past year	Missing	1	(Stm1-Stm3), (Stm6), (Stm8-Stm9), (Stm17)
16	Missing (lifetime use known)	Missing (lifetime use known)	Missing	125	(Stm1-Stm3), (Stm6), (Stm8), (Stm12-Stm15)
16	Missing (lifetime use imputed)	Missing (lifetime use imputed)	Missing	9	
16	Missing (lifetime use known)	Missing (lifetime use imputed)	Missing	0	
	Stimulants was logically assigned a recency category of "past month" or "past year but not past month use" based on the response to the Methamphetamines recency. There were no missing values.				
	Lifetime user, nothing missing			3,859	
	Imputed to lifetime nonuse			257	
	Lifetime nonuser, nothing missing			61,913	

[1]For one case, the missingness pattern was changed from pattern #6 to pattern #11 because the month of first use for stimulants indicated that methamphetamine use had to begin in the past year, making the respondent a past year methamphetamine user.

[2]For one case, the missingness pattern was changed from pattern #9 to pattern #11 because the month of first stimulants use indicated that methamphetamine user had to begin the past year, making the respondent a past year methamphetamine user.

Exhibit J.19 Health Insurance

Missingness Pattern				
#	Overall Health Insurance	Private Health Insurance	Number of Cases	Logical Constraints
1[1]	Missing		117	None
2	Missing	Missing	1,022	None
3		Missing	129	None

[1]This pattern only occurs if the response to the private health insurance question is "no". Obviously, if the response to the private health insurance question is "yes", the overall health insurance response would logically also be "yes."

This page intentionally left blank.

Section J.3

Exhibits Showing Missingness Patterns and the Portions of the Predictive Mean Vector Used in the Calculation of the Mahalanobis Distance (with Adjustments)

This page intentionally left blank.

Exhibit J.20 Portion of the Predictive Mean Vector for Cigarette Users

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	30-Day Frequency		
1	Past year	Missing	14	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)
2	Missing	Missing	154	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM
3	(Past month)	Missing	103	1. D 2. PM
4	Not past year		53	1. R3/(R3+R4)
5	Not past month		493	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. R3 = P(past 3 years but not past year use | lifetime use)
4. D = P(daily use | past month use)
5. PM = P(use on a given day in the past month | past month use)

Exhibit J.21 Portion of the Predictive Mean Vector for Cigar Users

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	30-Day Frequency		
1	Past year	Missing	14	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
2	Missing	Missing	117	1. R1 2. R2 3. R3 4. R1*PM
3	(Past month)	Missing	33	1. PM
4	Not past year		49	1. R3/(R3+R4)
5	Not past month		479	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)

[1] The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. R3 = P(past 3 years but not past year use | lifetime use)
4. PM = P(use on a given day in the past month | past month use)

Exhibit J.22 Portion of the Predictive Mean Vector for Smokeless Tobacco Users

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Number of Cases	Predictive Mean Vector[1]
1	(Past month)	(Past month)	Missing	Missing	32	1. DC 2. PMC 3. DS 4. PMS
2	(Past month)		Missing		43	1. DC 2. PMC
3		(Past month)		Missing	36	1. DS 2. PMS
4		Lifetime		Missing	20	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS
5	(Past month)	Lifetime	Missing	Missing	4	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
6	Lifetime		Missing		42	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC
7	Lifetime	(Past month)	Missing	Missing	1	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS
8		Past year		Missing	0	No cases
9	Past year		Missing		7	1. R1/(R1+R2) 2. RC1*DC/(RC1+RC2) 3. RC1*(1-DC)*PMC/(RC1+RC2)

Exhibit J.22 (continued)

Missingness Pattern						
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.	Num-ber of Cases	Predictive Mean Vector[1]
10	Lifetime	Lifetime	Missing	Missing	10	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS
11	Not past year				22	1. R3/(R3+R4)
12		Not past year			12	1. R3/(R3+R4)
13	Not past year	Not past year			2	1. R3/(R3+R4)
14	Not past month				192	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
15		Not past month			87	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
16	Not past month	Not past month			20	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)
17	Not past month	(Past month)		Missing	0	No cases
18	(Past month)	Not past month	Missing		1	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC
19	Not past month	Lifetime		Missing	2	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS

Exhibit J.22 (continued)

Missingness Pattern					Number of Cases	Predictive Mean Vector[1]
#	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.		
20	Lifetime	Not past month	Missing		0	No cases
21	Not past month	Not past year			0	No cases
22	Not past year	Not past month			0	No cases

[1]The predictive mean vector components are defined by the following:

1. R1 = P(past month smokeless tobacco use | lifetime smokeless tobacco use)
2. R2 = P(past year but not past month smokeless tobacco use | lifetime smokeless tobacco use)
3. R3 = P(past 3 years but not past year smokeless tobacco use | lifetime smokeless tobacco use)
4. 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 J.23 Portion of the Predictive Mean Vector for Alcohol Users

Missingness Pattern					Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Freq.	30-Day Freq.	30-Day Binge Drink		
1	(Past month)	Missing	Missing		90	1. PM 2. PY
2	(Past month)		Missing		402	1. PM
3	(Past month)	Missing			232	1. PY
4	(Past year but not past month)	Missing			333	1. PY
5	(Past month)			Missing	947	1. PMB
6	(Past month)		Missing	Missing	204	1. PM 2. PMB
7	(Past month)	Missing		Missing	45	1. PY 2. PMB
8	(Past month)	Missing	Missing	Missing	147	1. PM 2. PY 3. PMB
9	Past year		Missing	Missing	454	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)
10	Past year	Missing	Missing	Missing	52	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)
11	Lifetime	Missing	Missing	Missing	904	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB

[1] The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)
5. PMB = P(binge drinking on a given day in the past month | past month use)

Exhibit J.24 Portion of the Predictive Mean Vector for Marijuana Users

Missingness Pattern					
#	Recency	12-Month Frequency	30-Day Frequency	Number of Cases	Predictive Mean Vector[1]
1	(Past month)	Missing	Missing	60	1. PM 2. PY
2	(Past month)		Missing	54	1. PM
3	(Past month)	Missing		54	1. PY
4	(Past year but not past month)	Missing		103	1. PY
5	Past year		Missing	112	1. R1/(R1+R2) 2. R1*PM/(R1*R2)
6	Past year	Missing	Missing	80	1. R1/(R1+R2) 2. R1*PM/(R1*R2) 3. PY
7	Lifetime	Missing	Missing	585	1. R1 2. R2 3. R1*PM 3. (R1+R2)*PY

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Exhibit J.25 Portion of the Predictive Mean Vector for Cocaine Users

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
1	(Past month)		Missing		Missing		7	1. PM 2. PY
2	(Past month)				Missing		22	1. PM
3	(Past month)		Missing				4	1. PY
4	(Past year not past month)		Missing				21	1. PY
5	Past year				Missing		22	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year		Missing		Missing		30	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Missing		Missing		Missing		144	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
8	(Past month)	(Past month)		Missing		Missing	1	1. PM 2. PY
9	(Past month)	(Past month)				Missing	4	1. PM
10	(Past month)	(Past month)		Missing			0	No cases
11[2]	(Past year not missing)	(Past year but not past month)		Missing			3	1. PY
12	(Past month)	Past year				Missing	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
13	(Past month)	Past year		Missing		Missing	1	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY

Exhibit J.25 (continued)

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
14	(Past month)	Lifetime		Missing		Missing	4	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
15	(Past month)	(Past month)	Missing	Missing			0	No cases
16	(Past month)	(Past year but not past month)	Missing	Missing			1	1. PY
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			2	1. PY
18	(Past month)	(Past month)			Missing	Missing	1	1. PM
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	1	1. PM 2. PY
20	(Past month)	(Past month)	Missing		Missing	Missing	0	No cases
21	(Past month)	(Past month)		Missing	Missing	Missing	0	No cases
22	(Past month)	(Past month)	Missing	Missing	Missing		1	1. PM 2. PY
23	(Past month)	(Past year not past year)	Missing	Missing	Missing		1	1. PM 2. PY
24	(Past month)	(Past month)	Missing	Missing		Missing	0	No cases
25	(Past month)	(Past month)		Missing	Missing		0	No cases

Exhibit J.25 (continued)

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
26	(Past month)	(Past year not past month)		Missing	Missing		0	No cases
27	(Past month)	(Past month)	Missing			Missing	0	No cases
28	Past year	Past year			Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
29	Past year	Past year	Missing		Missing	Missing	0	No cases
30	Past year	Past year		Missing	Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
31	Past year	Past year	Missing	Missing	Missing	Missing	5	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
32	Past year	Lifetime		Missing	Missing	Missing	12	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
33	Past year	Lifetime	Missing	Missing	Missing	Missing	2	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
34	(Past month)	Lifetime		Missing	Missing	Missing	0	No cases
35	(Past month)	Lifetime	Missing	Missing	Missing	Missing	0	No cases

Exhibit J.25 (continued)

Missingness Pattern							Number of Cases	Predictive Mean Vector[1]
#	Coke Re-cency	Crack Re-cency	Coke 12-Mo. Freq.	Crack 12-Mo. Freq.	Coke 30-Day Freq.	Crack 30-Day Freq.		
36	Lifetime	Lifetime	Missing	Missing	Missing	Missing	37	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Note: Includes crack users, and cocaine users who were not crack users

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month cocaine use | lifetime cocaine use)
2. R2 = P(past year but not past month cocaine use | lifetime cocaine use)
3. PM = P(cocaine use on a given day in the past month | past month use of cocaine)
4. PY = P(cocaine use on a given day in the past year | past year use of cocaine)

[2]Due to a programming error, the three respondents fitting missingness pattern #11 were misclassified under missingness pattern #10. As a result, the donors assigned to these three respondents were all past month users, and the imputed 12-month frequency for crack might have been slightly affected.

Exhibit J.26 Portion of the Predictive Mean Vector for Heroin Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	0	No cases
2	(Past month)		Missing	1	1. PM
3	(Past month)	Missing		0	No cases
4	(Past year not past month)	Missing		2	1. PY
5	Past year		Missing	2	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	4	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Lifetime	Missing	Missing	31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Exhibit J.27 Portion of the Predictive Mean Vector for Hallucinogen Users

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
1		Lifetime				3	1. R1 2. R2
2			Lifetime			4	1. R1 2. R2
3		Lifetime	Lifetime			0	No cases
4	(Past month)			Missing	Missing	13	1. PM 2. PY
5	(Past month)				Missing	18	1. PM
6	(Past year)			Missing		91	1. PY
7	(Past month)	Lifetime			Missing	0	No cases
8	(Past month)		Lifetime		Missing	0	No cases
9	(Past month)	Lifetime	Lifetime		Missing	1	1. R1 2. R2 3. PM
10	(Past year)	Lifetime		Missing		0	No cases
11	(Past year)		Lifetime	Missing		0	No cases
12	(Past year)	Lifetime	Lifetime	Missing		0	No cases
13	(Past month)	Lifetime		Missing	Missing	0	No cases
14	(Past month)		Lifetime	Missing	Missing	1	1. R1 2. R2 3. PM 4. PY
15	(Past month)	Lifetime	Lifetime	Missing	Missing	0	No cases
16	Past year	(Not past month)	(Not past month)		Missing	18	1. R1/(R1+R2) 2. R1*PM/(R1+R2)

Exhibit J.27 (continued)

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
17	Past year	(Not past month)	(Not past month)	Missing	Missing	6	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
18	Past year	Past year	(Not past month)		Missing	7	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
19	Past year	(Not past month)	Past year		Missing	1	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
20	Past year	Past year	Past year		Missing	0	No cases
21	Past year	Lifetime	(Not past month)		Missing	6	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
22	Past year	(Not past month)	Lifetime		Missing	5	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
23	Past year	Lifetime	Lifetime		Missing	0	No cases
24	Past year	Past year	(Not past month)	Missing	Missing	0	No cases
25	Past year	(Not past month)	Past year	Missing	Missing	0	No cases
26	Past year	Past year	Past year	Missing	Missing	0	No cases
27	Past year	Lifetime	(Not past month)	Missing	Missing	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
28	Past year	(Not past month)	Lifetime	Missing	Missing	2	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
29	Past year	Lifetime	Lifetime	Missing	Missing	0	No cases

Exhibit J.27 (continued)

Missingness Pattern						Number of Cases	Predictive Mean Vector[1]
#	Hallucinogens Recency	LSD Recency	PCP Recency	Hallucinogens 12-Mo. Freq.	Hallucinogens 30-Day Freq.		
30	Lifetime	(Not past year)	(Not past year)	Missing	Missing	171	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
31	Lifetime	Lifetime	(Not past year)	Missing	Missing	129	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
32	Lifetime	(Not past year)	Lifetime	Missing	Missing	16	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY
33	Lifetime	Lifetime	Lifetime	Missing	Missing	22	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

Note: Hallucinogen users include users of LSD and PCP.

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Exhibit J.28 Portion of the Predictive Mean Vector for Inhalant Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency	30-Day Frequency		
1	(Past month)	Missing	Missing	10	1. PM 2. PY
2	(Past month)		Missing	12	1. PM
3	(Past month)	Missing		6	1. PY
4	(Past year not past month)	Missing		36	1. PY
5	Past year		Missing	35	1. R1/(R1+R2) 2. R1*PM/(R1+R2)
6	Past year	Missing	Missing	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY
7	Lifetime	Missing	Missing	436	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PM = P(use on a given day in the past month | past month use)
4. PY = P(use on a given day in the past year | past year use)

Exhibit J.29 Portion of the Predictive Mean Vector for Users of Pain Relievers, Tranquilizers, and Sedatives

Missingness Pattern			Number of Cases	Predictive Mean Vector[1]
#	Recency	12-Month Frequency		
1	(Past month)	Missing	PR: 77 TR: 11 SD: 4	1. PY
2	(Past year not past month)	Missing	PR: 89 TR: 21 SD: 5	1. PY
3	Past year		PR: 4 TR: 0 SD: 0	1. R1/(R1+R2)
4	Past year	Missing	PR: 9 TR: 1 SD: 0	1. R1/(R1+R2) 2. PY
5	Lifetime	Missing	PR: 597 TR: 214 SD: 87	1. R1 2. R2 3. (R1+R2)*PY

Note: The missingness patterns and predictive mean vectors for the pain relievers, tranquilizers, and sedatives modules were identical. When required, the identifiers "PR," "TR," and "SD" are used for pain relievers, tranquilizers, and sedatives, respectively.

[1]The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

Exhibit J.30 Portion of the Predictive Mean Vector for Stimulant Users

Missingness Pattern				Number of Cases	Predictive Mean Vector[1]
#	Stimulants Recency	Methamphet-amine Recency	12-Month Frequency		
1	(Past month)		Missing	100	1. PY
2	(Past year not past month)		Missing	215	1. PY
3	Past year			1	1. R1/(R1+R2)
4	Past year		Missing	11	1. R1/(R1+R2) 2. PY
5	Lifetime		Missing	180	1. R1 2. R2 3. (R1+R2)*PY
6		Lifetime		24	1. R1 2. R2
7	(Past month)	Lifetime	Missing	1	1. R1 2. R2 3. PY
8	(Past year not past month)	Lifetime	Missing	0	No cases
9	Past year	Lifetime		9	1. R1/(R1+R2)
10	Past year	Lifetime	Missing	0	No cases
11	(Past year)	Past year		0	No cases
12	(Past month)	Past year	Missing	0	No cases
13	(Past year not past month)	Past year	Missing	0	No cases
14	Past year	Past year		1	1. R1/(R1+R2)
15	Past year	Past year	Missing	1	1. R1/(R1+R2) 2. PY
16	Lifetime	Lifetime	Missing	134	1. R1 2. R2 3. (R1+R2)*PY

Note: Users of stimulants include users of methamphetamines.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(past month use | lifetime use)
2. R2 = P(past year but not past month use | lifetime use)
3. PY = P(use on a given day in the past year | past year use)

Exhibit J.31 Health Insurance

Missingness Pattern				
#	Overall Health Insurance	Private Health Insurance	Number of Cases	Predictive Mean Vector
1	Missing		117	1. Overall health insurance
2	Missing	Missing	1,022	1. Overall health insurance 2. Private health insurance
3		Missing	129	1. Private health insurance

Appendix K

1999 National Household Survey on Drug Abuse Specifications for Programming

This page intentionally left blank.

The 1999 National Household Survey on Drug Abuse Specifications for Programming is available on the web at http://www.drugabusestatistics.samhsa.gov/.