METHODS FOR HANDLING MISSING ITEM VALUES IN REGRESSION MODELS USING THE NATIONAL SURVEY ON DRUG USE AND HEALTH (NSDUH)

NSDUH METHODOLOGICAL REPORT

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland

September 2018

This page intentionally left blank

METHODS FOR HANDLING MISSING ITEM VALUES IN REGRESSION MODELS USING THE NATIONAL SURVEY ON DRUG USE AND HEALTH (NSDUH)

NSDUH METHODOLOGICAL REPORT

For questions about this report, available from https://www.samhsa.gov/data/, please e-mail Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

September 2018

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2018). *Methods for Handling Missing Item Values in Regression Models Using the National Survey on Drug Use and Health (NSDUH): NSDUH Methodological Report*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

This page intentionally left blank

Table of Contents

Section **Page**

Table of Contents (continued)

Section **Page**

Appendix

List of Tables

Table **Page**

List of Tables (continued)

Table **Page**

List of Tables (continued)

Table **Page**

List of Exhibits

Exhibit **Page**

This page intentionally left blank

1. Introduction

The purpose of this report is to guide analysts interested in fitting regression models using data from the National Survey on Drug Use and Health (NSDUH) by providing them with methods for handling missing item values in regression analyses (MIVRA). In addition, this report can serve as general guidance for analysts interested in MIVRA methods for complex surveys other than NSDUH. This report focuses on missingness in independent variables for regression analyses. Even though the extent of missing data for an individual item is typically very low on NSDUH, when multiple variables are being used in an analysis (such as when multiple independent variables are used in a regression analysis), the number of cases with at least one variable with missing data has the potential to increase. The report includes a theoretical review of existing MIVRA methods, a simulation study that evaluates several of the more promising methods using existing NSDUH datasets, and a final chapter where the results of both the theoretical review and the simulation study are synthesized into guidance for analysts via decision trees. More specific descriptions of the chapters of this report follow. A list of contributors to this report appears before the appendices.

In Chapter 2, previous NSDUH studies involving regression models are examined to determine the extent and patterns of item nonresponse typically faced by NSDUH analysts. This review of analyses provides the background for assessing the MIVRA methods described in later chapters. A secondary goal of this review is to select a few regression analyses for the simulation study described in Chapter 4.

Chapter 3 describes existing MIVRA methods and summarizes the available literature on their intended use and typical performance within complex sample surveys. The methods reviewed include listwise deletion (complete case analysis), weighting methods, pairwise deletion, imputation, (pseudo-)maximum likelihood, and the inclusion of an extra term in the model to denote a missing item value. The methods deemed most promising (for use with NSDUH data) based on the literature are used in the simulation study described in Chapter 4.

Chapter 4 describes the simulation study that evaluated the most promising MIVRA methods identified in Chapter 3 using a few NSDUH regression analyses selected in Chapter 2. The six MIVRA methods were applied to six regression models and evaluated (mainly with respect to bias, variance, and accuracy of variance estimation). The six MIVRA methods selected to be used in the simulation study include listwise deletion (the simplest method); a variant on listwise deletion where the complete observations are reweighted; two hot-deck imputation methods; and pseudo-maximum-likelihood estimation methods that are available in two software packages (Mplus® and Latent GOLD®).

Finally, Chapter 5 provides a summary of the recommended MIVRA methods from the six methods selected for inclusion in the simulation to be applied to NSDUH data, based on both the theory in the literature review and the results of the simulation study. The factors affecting the choice of method include the extent of missing item values, the ease of implementation of the MIVRA method, and the anticipated effects on the bias and variance of coefficient estimates. Analysts who use other sources of data may find it useful to conduct simulation experiments like those described in Chapter 4 using the missingness rates and properties from their own complex survey data when there are potentially unacceptable amounts of missing item values. Readers interested primarily in receiving guidance (and less so in the technical reasoning behind that guidance) might read only Chapter 5, along with the description of the recommended MIVRA method(s) in Chapter 3 and the sample code in the appendices.

This report focuses on MIVRA methods that could be used for item-level missingness in independent variables of a regression analysis (particularly, logistic regression) but does not address the following topics in detail.

- *Unit nonresponse.* An important assumption in this report is that unit (whole-record) nonresponse in NSDUH had been handled in a nearly (asymptotically) unbiased fashion by reweighting. See the person-level sampling weight calibration report in the 2015 NSDUH methodological resource book (Center for Behavioral Health Statistics and Quality, 2017a) for full descriptions of the final analysis weights.
- *Missingness in variables that have already been treated by the annual imputation procedures.* For this report, a decision was made to consider most NSDUH variables that already have imputed values as part of the annual imputation treatment as though they have no missing values. The rationale is that the NSDUH imputation treatment in most cases is fairly sophisticated and is expected to correct for nonresponse bias reasonably well, though an exception and its reversal are discussed in Section 2.2.2. Moreover, the underestimation of variance due to imputed values being treated as actual responses is expected to be low. See Chapter 6 of the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* (Center for Behavioral Health Statistics and Quality, 2017b) for justification of these claims.
- *Missingness in the dependent variable.* Respondents with a missing dependent variable tend to provide little to no information that can be used in the modeling, even if none of the independent variables are missing (Little, 1992). However, it is sometimes possible to reweight for nonresponse when a record is listwise deleted because the dependent variable is missing and the deletion is (partly) a function of the dependent variable's value. This method is described in Section 5.4. Often, as is the case here, records with a missing dependent variable are simply dropped from the analysis. This can be justified when the probability of the variable being missing is a function of the independent variables in the model that are never missing (see Section 3.2.2).
- *Missingness in the subpopulation indicator.* NSDUH analyses tend to be of subpopulations. In fact, every regression analysis reviewed in Chapter 2 was of a subpopulation based on the age of the respondent, lifetime or recent use of certain drugs, etc. In this report, the assumption is that either the subpopulation indicator is never missing or its missingness is a function of the independent variables in the model that are never missing (see Section 3.2.2). Based on the review in Chapter 2, subpopulation indicators in NSDUH analyses tend to have undergone the annual imputation procedures or to have few missing values.
- *Missingness that should be treated as a valid response.* It is up to the analyst to decide whether "missing" is a valid response or whether it substitutes for a response that ideally should have been provided. For example, "Don't Know" responses are often treated as valid for the availability-of-drug variables, in that a respondent may genuinely not know how difficult or easy it would be to obtain heroin if he or she wanted some. In this instance, the "Don't Know" response category may be of analytic interest and would not be considered a missing value.

2. Extent of Missingness in NSDUH Analyses

2.1 Introduction

In general, item response rates on the National Survey on Drug Use and Health (NSDUH) tend to be consistently high across time and different variables; see the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* (Center for Behavioral Health Statistics and Quality, 2017b) for details. For example, of the 181 variables in the 2014 survey that underwent the standard complex imputation treatment, only 12 had more than 5 percent item nonresponse before imputation.[1]

Even with a high item response rate for individual variables, the percentage of respondents dropped from a regression model using the popular method of excluding all records with any missing item values—called "listwise deletion" or "complete case analysis"—can still be high when the model contains many covariates. For example, consider a model with 10 independent variables. If each independent variable has an item nonresponse rate of 2 percent, the percentage of respondents with a missing value for one or more independent variables can theoretically range from 2 percent to 20 percent. The more independent variables there are, the greater the potential for more erosion in the model's sample size and representativeness when respondents with any missing item values are dropped.

In this chapter, the extent of missing item values in NSDUH is identified by assessing a selection of NSDUH studies that use logistic regression analyses. Logistic regression is a frequent type of analysis used by the Center for Behavioral Health Statistics and Quality when analyzing NSDUH data. Therefore, the following sections focus on assessment of item missingness and particularly on missingness in NSDUH studies across independent variables in a logistic regression analysis.

2.2 Methods for Assessing Item Missingness in NSDUH Analyses

For purposes of this study, the extent of missingness in NSDUH analyses was assessed in two steps. The first step was a preliminary screening of 16 recent NSDUH analytic studies (Section 2.2.1). These studies are identified throughout this report by the NSDUH analytic study codes[2] used internally by the Center for Behavioral Health Statistics and Quality. Simple statistics related to missingness were collected for each of these studies. For the second step, 3 of the 16

[1] Note that, as mentioned in Chapter 1, not all variables undergo the standard complex imputation treatment. Item response rates are not tabulated for the variables that do not undergo imputation, so overall knowledge of item response rates is somewhat limited. However, because many members of the set of variables that do undergo imputation store sensitive information, such as illegal use of drugs, it is expected that, in general, item response rates of variables that do not undergo imputation are not appreciably lower than item response rates of variables that do undergo imputation.

[2] The 16 NSDUH analytic study codes are C8, C10, I1, I2, P4, T1, T2, N1, N4, N14, N15, N18, N19, PR2, PR5a, and PR7.

studies were selected for a more detailed analysis of missingness (Section 2.2.2).[3] In this second step, most ad hoc imputation procedures were undone (creating a new dataset with more missing item values)[4] and the missingness statistics were more detailed. The missingness properties examined in these three studies were used in a simulation experiment described in Chapter 4.

2.2.1 Preliminary Screening for All 16 Studies

All 16 analytic studies examined in the preliminary screening step involved at least one regression model. In total, 55 models were examined, and each was the final version used in the study. Earlier models that may have included more or different covariates were not considered. Statistics describing missingness rates and potential for bias in coefficient estimates due to missingness were computed. For these statistics, all variables that underwent the standard complex NSDUH imputation treatment were assumed to have no missing item values.[5]

The following statistics were computed for each of these 55 models.

- *The percentage of respondents in the subpopulation of interest with a missing value for the regression model's dependent variable.*
- *Out of those respondents in the subpopulation of interest with a nonmissing value for the dependent variable, the percentage of respondents with a missing value for one or more independent variables.* Respondents with a missing dependent variable tend to provide little to no information that can be used in the modeling, even if none of the independent variables are missing (Little, 1992). Thus, the *deletion rate* is defined as the *fraction of cases in the subpopulation of interest with a nonmissing dependent variable that has at least one missing independent variable*.
- *The weighted distribution of the dependent variable using all respondents in the subpopulation of interest with a nonmissing value for the dependent variable, and the weighted distribution of the dependent variable using all respondents in the subpopulation of interest with nonmissing values for both the dependent variable and all independent variables.* These two statistics give some idea of the bias that might be introduced when only "complete" cases (i.e., those with no missing data) are used in a regression analysis. For 53 of the 55 models, the dependent variable is dichotomous, so the weighted distribution is called the "prevalence" (i.e., the weighted percentage of affirmative responses).

2.2.2 Detailed Analysis of Missingness for Selected Studies

Three of the 16 studies were selected for a more detailed analysis after the preliminary screening was complete. Although it would have been a better assessment of the full extent of the missingness to undo the additional imputations before assessing missingness for all 16

---

[3] The NSDUH analytic study codes for the three selected studies are N4, N14, and N19.

[4] Section 2.2.2 explains the distinction between the standard complex imputation treatment, which was not undone, and the ad hoc imputation procedures, which were undone, in the detailed analysis of missingness.

[5] As pointed out in Chapter 1, the standard complex NSDUH imputation methodology is expected to correct for nonresponse bias reasonably well, and the underestimation of variance due to imputed values being treated as actual responses is expected to be low. An additional NSDUH imputation, called "zero-fill," was treated differently, as will be shown.

studies, this was not feasible for the preliminary screening described because of the large number of studies and models involved. The reasons for choosing these three studies were based on their missingness rates and whether the subpopulation of interest was for youths or adults. The details of this selection are provided in Section 2.3.1. Moreover, each of the three studies had two models that were included in this analysis, for a total of six models. The analysis datasets were examined closely in two stages and reprocessed as described below, and a more complex and informative set of statistics was produced for these three studies.

2.2.2.1 Assessment of Missingness in the Subpopulation Indicator

Each of the three selected studies analyzed subpopulations, which are typically identified by a 0/1 subpopulation indicator variable used to indicate whether (=1) or not (=0) the respondent is a member of the subpopulation of interest. Sometimes these indicator variables have missing values themselves. This usually happens in the presence of filter questions, as described in Eckman et al. (2014). Filter questions are commonly used in surveys, including NSDUH, to decide which respondents are presented with a set of follow-up questions. For example, in NSDUH, only respondents reporting lifetime use of alcohol (i.e., the filter question) are presented with follow-up questions about how recently they used alcohol and their age when they first used alcohol. Some variables associated with the responses to filter questions do not undergo complex NSDUH imputation. Therefore, membership of some respondents in a particular model's subpopulation of interest is unknown. For each of the six models, missingness in the subpopulation indicator was assessed.[6]

2.2.2.2 Additional Treatment of Missing Item Values in NSDUH Analyses

In all studies, variables that were created using NSDUH's standard complex imputation process were used and were assumed to have no missing data. However, not all variables in NSDUH and in these analytic studies undergo the standard complex imputation treatment. A few, including the substance dependence and abuse variables and some of the adult mental health variables, undergo a separate, standard annual treatment where their missing values are replaced with zeroes (a process described as "zero-fill" imputation). Several other variables underwent "ad hoc" imputation that was specific to a particular analysis. These ad hoc methods included (1) weighted sequential hot-deck (WSHD) imputation within predetermined imputation cells formed by cross-classifying race/ethnicity (Hispanic/Latino, non-Hispanic/Latino white, non-Hispanic/Latino black/African American, and non-Hispanic/Latino other) and gender (when applicable) and then sorting by age;[7] and (2) zero-fill imputation.

For the detailed analyses of missingness, an attempt was made to undo all imputation procedures other than those that used NSDUH's complex imputation method.[8] Most of the ad hoc imputation methods used in these three selected studies were undone because the purpose of

---

[6] This report touches only briefly on methods for handling missingness in the dependent variable and/or the subpopulation indicator. See Section 3.6.

[7] This method is described in more detail in Chapter 3. It was also used in the simulation experiment described in Chapters 4 and 5.

[8] The standard complex imputation treatment was not undone, because in the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* (Center for Behavioral Health Statistics and Quality, 2017b), it was determined that these procedures work reasonably well.

this analysis is to assess the missingness rates before the missing item values could be treated with additional imputation steps and to standardize this treatment across all NSDUH analyses. For some of the zero-filled variables created specifically for these studies, the zero-fill imputation method could be undone relatively easily. For example, the standard zero-fill imputation methods applied to the substance dependence and abuse variables and some of the adult mental health variables were undone relatively easily because alternate versions of these variables were already created for other studies (Center for Behavioral Health Statistics and Quality, 2017b).[9] However, the two N4 models[10] used some zero-filled variables where replacement of values with missing values would have been difficult. Thus, a limited number of zero-filled variables remained for this assessment.

After reprocessing the datasets as described above, for each of the six models, the following statistics were determined.

2.2.2.3 Record-Level Statistics Involving the Model (Table 2.3, Tables 4.3 to 4.8, and Appendix A)

- *The percentage of respondents whose membership in the subpopulation of interest is unknown.*
- *The percentage of respondents known to be in the subpopulation of interest with missing values for the dependent variable.*
- *The number and percentage of respondents in each missingness pattern.* After keeping only the observations (1) known to be in the subpopulation of interest, and (2) those with a nonmissing value for the dependent variable, tables were created that show the number and percentage of observations with each pattern of missingness in the independent variables to understand the nature and extent of these missingness patterns. These tables appear in Section 4.3, where their function in the simulation experiment is described.
- *The deletion rate.* As stated in Section 2.2.1, this is the percentage of respondents known to be in the subpopulation of interest with a nonmissing dependent variable that have at least one independent variable that has missing values.

2.2.2.4 Variable-Level Statistics Involving the Model (Tables 2.4 to 2.9)

- *Imputation treatment*. This had six possible values.
  - *No missing values*. These variables had no missing values by design. For example, if the respondent fails to answer the gender question, he or she is treated as a unit nonrespondent (and unit nonresponse is handled with reweighting). The age variable is edited in such a way that there are no missing values. Also, the geographic variables used in sampling are known for all people in the target population.

[9] In these other studies, it was determined that the zero-fill imputation method induced a noticeable negative bias for some of the substance dependence and abuse variables, but the negative bias was not as noticeable for the adult mental health variables.

[10] These variables concerned past month serious psychological distress among women aged 18 to 44. More details can be found in Table 2.2.

   - *Complex NSDUH imputation*. These variables underwent the standard complex NSDUH imputation treatment and were treated as if they had no missing values.
   - *Ad hoc WSHD*. These variables underwent WSHD imputation within predetermined imputation cells when used in the specific model for a given analysis.
   - *Standard zero-fill imputation*. For these variables, missing values were replaced by zeroes (or negative responses) as part of the standard variable creation process.
   - *Ad hoc zero-fill imputation*. These variables were zero-filled when used in the specific model.
   - *No imputation*. These variables had missing values when the model was fit during the analytic study and for the analyses presented later in this report.

- *Across respondents in the subpopulation of interest, the number and percentage of each independent variable with a missing value*. This is the variable-level equivalent of the deletion rate.
- *Details on the dependent variable*. This includes the number of levels and the description of the levels. All six dependent variables were categorical.

2.3 Results of the Item Missingness Assessment in NSDUH Studies

Section 2.3.1 discusses the results of the preliminary screening of missingness for all 16 studies. Section 2.3.2 reports the results of the detailed analysis of missingness for the three studies and six models selected for closer examination.

2.3.1 Results of Preliminary Screening of Missingness for All 16 Studies

Table 2.1 shows some basic statistics on missingness for the 16 studies. A more detailed version of this table is available in Appendix B. Because some of the variables retained their imputed values for these analyses, the true missingness rate could be higher. The results can be summarized as follows.

- None of the models had a high percentage of respondents with a missing dependent variable (all less than 1.5 percent).
- Some of the studies had higher deletion rates and therefore the potential for nonresponse bias. Fifteen of the 55 models had deletion rates of at least 10 percent. A few of the studies (e.g., P4, N15, and PR5a) used mostly variables that underwent NSDUH's complex imputation treatment and are considered for the purpose of this effort to have no missing values. Other studies, such as I1, N14, and PR2, used several variables that did not undergo complex NSDUH imputation, and the deletion rate was 15 percent or higher.

Table 2.1 Summary of Missingness Statistics for 16 NSDUH Studies

Study[1]	Number of Models	Respondents with Dependent Variable Missing (%)	Deletion Rate (%)	Prevalence of Dependent Variable Using All Records Where Dependent Variable Is Not Missing (%)	Prevalence of Dependent Variable Using Only Records with No Missing Item Values (%)
C8	2	None/Negligible[2]	2.00-3.00	Negligible differences[3]	
C10	3	None/Negligible	5.50-6.00	5.75	5.66-5.68
I1	7	None/Negligible	13.00-15.00 for 4 models involving youths; 2.00-3.00 for 3 models involving young adults	Differences up to 0.43 percentage points	
I2	1	None/Negligible	0.60	Negligible differences	
P4	4	None/Negligible	None/Negligible	N/A	
T1	3	None/Negligible	11.00-14.00	Differences up to 0.39 percentage points	
T2	3	None/Negligible	2.50-3.00	Negligible differences[3]	
N1	4	None/Negligible	1.10-1.30	Differences up to 0.23 percentage points	
N4	2	None/Negligible	3.00-4.00	Differences up to 0.19 percentage points	
N14	2	Up to 1.12	14.00-15.00	*Y* has 5 levels; some noticeable differences	
N15	3	None/Negligible	None/Negligible	Negligible differences	
N18	3	Up to 0.43	0.86	Differences up to 0.12 percentage points	
N19	4	None/Negligible	10.00	Differences up to 0.70 percentage points	
PR2	2	None/Negligible	26.00-27.00	Differences up to 2.10 percentage points	
PR5a	11	None/Negligible	None/Negligible	Negligible differences	
PR7	1	None/Negligible	4.77	2.60	2.72

N/A = not applicable.
Note: These statistics are described in Section 2.2.1.
[1] These 16 studies are identified by the NSDUH analytic study codes used internally by SAMHSA.
[2] Percent missing is 0.10 or less.
[3] Absolute differences in percentages are 0.10 or less.

The three studies that were selected for closer examination were chosen based on the results in this table. N14 (a study on helpfulness of treatment among adolescents with a major depressive episode in the previous year) and N19 (a study on adolescent use and abuse of pain relievers) were chosen because their deletion rates were high,[11] and the last two columns of Table 2.1 suggest the potential for nonresponse bias. N14 presents an additional challenge because, for some respondents, membership in the subpopulation of interest is uncertain. Unlike N14 and N19, N4 (a study on serious psychological distress among adult pregnant women) was also chosen because it involved respondents aged 18 or older, whereas other selected studies were on youths, and missingness rates in NSDUH tend to differ for youths and adults. The three selected studies and six models are described in some detail in Table 2.2. The independent variables used in each model are listed in Tables 2.4 through 2.9 in Section 2.3.2.

[11] Despite its high deletion rate, PR2 was not chosen, because the model selection for this study had not been finalized at the time that this project began.

Table 2.2 Studies and Models Selected for Detailed Missingness Analysis

Study	Years	Model	Subpopulation of Interest	Dependent Variable	Sample Size[1]
N4	2008-2012	1	Women aged 18-44	Past month SPD	93,100
		2	Women aged 18-44 with past month SPD	Past year mental health treatment	7,600
N14	2006-2010	1	Adolescents aged 12-17 with past year MDE who reported receiving counseling in the past year	Helpfulness of counseling (5 levels)	3,300
		2	Adolescents aged 12-17 with past year MDE who reported taking a prescribed medication for depression in the past year	Helpfulness of medication (5 levels)	1,500
N19[2]	2008-2012	1	Adolescents aged 12-17	Past year use of pain relievers	112,600
		2	Adolescents reporting past year use of pain relievers	Past year pain reliever disorder	7,100

MDE = major depressive episode; SPD = serious psychological distress.

[1] Reported sample sizes are rounded to the nearest 100.

[2] Although this study involved four models (see Table 2.1), only two are listed here because three of the models differ only in their use of interaction terms. Interaction terms do not affect the deletion rate. Therefore, there was little benefit from considering these three similar models separately.

2.3.2 Results of Detailed Analysis of Missingness for the Three Selected Studies

This section reports the results of the detailed analysis of missingness for the studies and models listed in Table 2.2.

Table 2.3 lists the deletion rates for each of the six models. Note that the deletion rate for each study, after reprocessing, increased. This is in part due to undoing the standard zero-fill imputation procedures applied to the substance dependence and abuse variables, which appear in all six models. The increases observed in the detailed analysis suggest that the deletion rates reported in the preliminary screening (Section 2.3.1) are likely to be underestimates for many models. Note also that the deletion rates vary widely among the six models. The N14 and N19 studies included adolescents only, and item nonresponse is generally higher among adolescents.

Table 2.3 Deletion Rates of the Six Models Selected for the Detailed Missingness Analysis

Model	Deletion Rate before Reprocessing (%)	Deletion Rate after Reprocessing (%)	Relative Increase (%)
N4 Model 1	3.64	4.37	20.05
N4 Model 2	3.16	4.71	49.05
N14 Model 1	13.79	15.84	14.87
N14 Model 2	15.14	17.15	13.28
N19 Model 1	10.33	13.14	27.20
N19 Model 2	10.28	12.58	22.37

Note: The methods used in this analysis are described in Section 2.2.2.

Tables 2.4 through 2.9 report variable-level missingness for each of the six models. More details on the precise handling of each of these variables are available in Appendix A.

Descriptions of the entries in the "Imputation Treatment" column are provided in Section 2.2.2. Note that, despite what was mentioned in Section 2.1 about erosion of the sample size due to a large number of independent variables, for these six models at least, only a few variables are driving the deletion rates. For the N14 models, the number of mental health visits and grades in school had relatively low item response rates, and for the N19 models, grades in school and family support had relatively low item response rates.

Table 2.4 Missingness Statistics for Model Variables, N4 Model 1

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variables			
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Dependent Variable			
Past Month Serious Psychological Distress	Standard zero-fill imputation; undone	426	0.46
Independent Variables			
Pregnancy Status	No imputation	26	0.03
Age Recode	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Marital Status (3 levels instead of 4)	Complex NSDUH imputation	N/A	N/A
Education Level	Complex NSDUH imputation	N/A	N/A
Employment Status	Complex NSDUH imputation	N/A	N/A
Federal Poverty Level	No imputation	2,535	2.73
Rapid Repeat Birth	Ad hoc zero-fill imputation; too complex to undo	N/A	N/A
Number of Biological Children in Household	Ad hoc zero-fill imputation; too complex to undo	N/A	N/A
Health Insurance	Complex NSDUH imputation	N/A	N/A
Health Status	No imputation	14	0.02
Health Problems	Ad hoc zero-fill imputation, in part; undone	819	0.88
Past Month Cigarette Use	Complex NSDUH imputation	N/A	N/A
Past Year Alcohol Use Disorder	Ad hoc zero-fill imputation; undone	322	0.35
Past Year Illicit Drug Use (+ Disorder)	Ad hoc zero-fill imputation; undone	418	0.45

N/A = not applicable.

Table 2.5 Missingness Statistics for Model Variables, N4 Model 2

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variables			
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Past Month Serious Psychological Distress	Standard zero-fill imputation; undone	426	0.46

Table 2.5 Missingness Statistics for Model Variables, N4 Model 2 (continued)

Variable	Imputation Treatment	Number Missing	Percent Missing
Dependent Variable			
Past Year Mental Health Treatment	No imputation	26	0.34
Independent Variables			
Pregnancy Status	No imputation	2	0.03
Age Recode	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Marital Status (3 levels instead of 4)	Complex NSDUH imputation	N/A	N/A
Education Level	Complex NSDUH imputation	N/A	N/A
Employment Status	Complex NSDUH imputation	N/A	N/A
Federal Poverty Level	No imputation	178	2.35
Rapid Repeat Birth	Ad hoc zero-fill imputation; too complex to undo	N/A	N/A
Number of Biological Children in Household	Ad hoc zero-fill imputation; too complex to undo	N/A	N/A
Health Insurance	Complex NSDUH imputation	N/A	N/A
Health Status	No imputation	3	0.04
Health Problems	Ad hoc zero-fill imputation, in part; undone	74	0.98
Past Month Cigarette Use	Complex NSDUH imputation	N/A	N/A
Past Year Alcohol Use Disorder	Standard zero-fill imputation; undone	38	0.50
Past Year Illicit Drug Use (+ Disorder)	Standard zero-fill imputation; undone	70	0.92
Had Depression in Lifetime	No imputation	57	0.75
Had Anxiety in Lifetime	No imputation	57	0.75

N/A = not applicable.

Table 2.6 Missingness Statistics for Model Variables, N14 Model 1

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variables[1]			
Age	No missing values	N/A	N/A
Past Year Major Depressive Episode	No imputation	2,293	2.05
Past Year Counseling	No imputation	246	0.22
Dependent Variable			
Helpfulness of Counseling	No imputation	37	1.12
Independent Variables			
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Family Income	Complex NSDUH imputation	N/A	N/A
Health Insurance	Complex NSDUH imputation	N/A	N/A
Rural/Urban	No missing values	N/A	N/A
Number of Delinquent Behaviors	Ad hoc zero-fill imputation; undone	10	0.31

Table 2.6 Missingness Statistics for Model Variables, N14 Model 1 (continued)

Variable	Imputation Treatment	Number Missing	Percent Missing
Grades	WSHD; undone	210	6.42
Family Encouragement	Ad hoc zero-fill imputation; undone	4	0.12
Religious Services	WSHD; undone	23	0.70
Severe Impairment	No imputation	8	0.24
Past Year Substance Use Disorder	Standard zero-fill imputation; undone	79	2.42
Number of Mental Health Visits	WSHD; undone	239	7.31
Past Year Mental Health Medications	No imputation	3	0.09

N/A = not applicable; WSHD = weighted sequential hot-deck imputation within predetermined imputation cells.
[1] A total of 303 (0.27 percent) respondents in the dataset had a missing value for the subpopulation indicator (i.e., they were missing for both past year major depressive episode [MDE] and past year counseling, missing for past year MDE and positive for past year counseling, or positive for past year MDE and missing for past year counseling).

Table 2.7 Missingness Statistics for Model Variables, N14 Model 2

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variables[1]			
Age	No missing values	N/A	N/A
Past Year Major Depressive Episode	No imputation	2,293	2.05
Past Year Medication	No imputation	1,966	1.76
Dependent Variable			
Helpfulness of Medication	No imputation	6	0.39
Independent Variables			
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Family Income	Complex NSDUH imputation	N/A	N/A
Health Insurance	Complex NSDUH imputation	N/A	N/A
Rural/Urban	No missing values	N/A	N/A
Number of Delinquent Behaviors	Ad hoc zero-fill imputation; undone	5	0.32
Grades	WSHD; undone	89	5.78
Family Encouragement	Ad hoc zero-fill imputation; undone	3	0.19
Religious Services	WSHD; undone	11	0.71
Severe Impairment	No imputation	4	0.26
Past Year Substance Use Disorder	Standard zero-fill imputation; undone	36	2.34
Number of Mental Health Visits	WSHD; undone	139	9.03

N/A = not applicable; WSHD = weighted sequential hot-deck imputation within predetermined imputation cells.
[1] A total of 211 (0.19 percent) respondents in the dataset had a missing value for the subpopulation indicator (i.e., they were missing for both past year major depressive episode [MDE] and past year medication, missing for past year MDE and positive for past year medication, or positive for past year MDE and missing for past year medication).

Table 2.8 Missingness Statistics for Model Variables, N19 Model 1

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variable			
Age	No missing values	N/A	N/A
Dependent Variable			
Past Year Misuse of Pain Relievers	Complex NSDUH imputation	N/A	N/A
Independent Variables			
Past Year Major Depressive Episode	WSHD; undone	2,414	2.14
Family Support	Ad hoc zero-fill imputation; undone	8,253	7.33
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Family Income	Complex NSDUH imputation	N/A	N/A
Rural/Urban	No missing values	N/A	N/A
Alcohol Use Disorder	Standard zero-fill imputation; undone	1,030	0.91
Illicit Drug Use Disorder (excluding pain relievers)	Standard zero-fill imputation; undone	1,382	1.71
Number of Delinquent Behaviors	Ad hoc zero-fill imputation; undone	992	0.88
Grades	WSHD; undone	8,382	7.44
Religious Services	WSHD; undone	2,769	2.46

N/A = not applicable; WSHD = weighted sequential hot-deck imputation within predetermined imputation cells.

Table 2.9 Missingness Statistics for Model Variables, N19 Model 2

Variable	Imputation Treatment	Number Missing	Percent Missing
Subpopulation Variable			
Age	No missing values	N/A	N/A
Past Year Misuse of Pain Relievers	Complex NSDUH imputation	N/A	N/A
Dependent Variable			
Past Year Pain Reliever Disorder	Standard zero-fill imputation; undone	788	12.58
Independent Variables			
Past Year Major Depressive Episode	WSHD; undone	130	2.06
Family Support	Ad hoc zero-fill imputation; undone	375	5.96
Age	No missing values	N/A	N/A
Gender	No missing values	N/A	N/A
Race/Ethnicity	Complex NSDUH imputation	N/A	N/A
Family Income	Complex NSDUH imputation	N/A	N/A
Rural/Urban	No missing values	N/A	N/A
Alcohol Use Disorder	Standard zero-fill imputation; undone	64	1.02
Illicit Drug Use Disorder (excluding pain relievers)	Standard zero-fill imputation; undone	177	2.81
Number of Delinquent Behaviors	Ad hoc zero-fill imputation; undone	56	0.89
Grades	WSHD; undone	350	5.56
Religious Services	WSHD; undone	105	1.67

N/A = not applicable; WSHD = weighted sequential hot-deck imputation within predetermined imputation cells.

2.4 Conclusions

The missingness assessment of the 16 NSDUH analytic studies, including the detailed examination of the 3 selected studies, leads to the following conclusions.

- *Missingness tends to be more prevalent for studies involving respondents aged 12 to 17.* All studies with deletion rates higher than 10 percent involve youths: I1, T1, N14, N19, and PR2.
- *The substance dependence and abuse variables and the mental health variables, which undergo zero-fill imputation, are frequently used.* All three studies selected for closer examination employed some variables from this set. The project staff who conducted the other studies confirmed that most analyses involved these variables.
- *Ad hoc zero-fill imputation appears to be frequently used.* All three studies selected for closer examination employed ad hoc zero-fill imputation. The most noteworthy example was study N19, in which the family support variable had no missing values in the models used in that study but had 6 to 7 percent missingness when alternate versions were created for the current NSDUH study.
- *Undoing zero-fill imputation tends to increase the deletion rate.* For the six models that underwent closer examination, Table 2.3 shows that the deletion rates reported in the preliminary screening analysis were lower than the deletion rates reported after zero-fill imputation was undone.
- *Missingness tends to be driven by a few variables.* For all three studies that underwent close examination, there were one or two variables that had more missingness than others. Both studies involving respondents aged 12 to 17 (i.e., N14 and N19) used the respondent's grades as an independent variable; this variable seems to have relatively high nonresponse.
- *Sometimes deletion rates exceed 10 percent*. Although most NSDUH variables undergo a complex imputation treatment, those variables that do not undergo this treatment are often used in logistic regression analyses of NSDUH data. Therefore, the use of these variables in analyses leads to studies that have fairly high deletion rates, which introduces the potential for nonresponse bias.

3. Available Methods for Handling Missing Item Values in Regression Analyses of Complex Survey Data

3.1 Introduction

Based on the analysis described in Chapter 2, it was determined that the rates of missing item values in regression analyses for the National Survey on Drug Use and Health (NSDUH) are high enough to consider alternatives to "complete case analyses" that use only the cases with no missing data. Section 3.2 presents the results from a literature review of several candidate methods for handling missing item values in regression analyses (MIVRA). Section 3.3 lists the MIVRA methods that were most promising based on the literature review and the availability of the methods in the software. These methods were assessed in more detail in the simulation study described in Chapter 4.

3.2 Literature Review on MIVRA Methods for Complex Survey Data

The subject of missing item values in statistical analyses has garnered considerable attention in the past 30 years, spurred by the initial publication of Little and Rubin's *Statistical Analysis of Missing Data* (1987; 2002). Many books and articles summarize "best practices" for handling analyses with missing item values, including Pigott (2001), Schafer and Graham (2002), Allison (2002), and Horton and Kleinman (2007). However, most of these publications do not address the impact of the complex design often used in survey samples and sometimes concentrate wholly on clinical trials. This includes the recent and otherwise extensive report by the National Research Council (2010).

Notable exceptions to the general absence of discussion related to complex survey data are Kalton and Kaspryzyk (1982) and, more recently, Kim and Shao (2013). The analyses in the former, unlike many other investigations of missing item values in surveys, go beyond exploring the impact of missing item values on the estimation of population means and totals by also focusing on population distributions and covariances. The treatment of variance measures for the estimates discussed in this pioneering work is, however, understandably limited.

Although Kim and Shao's (2013) text addresses statistical analyses with missing item values in general, it provides a useful chapter on applications to survey samples. The authors prefer a method for handling missing item values, fractional imputation, which was not supported by any statistical software at the time the analysis detailed in this report was undertaken.[12] Thus, fractional imputation is not discussed here. Instead, brief treatments of the following MIVRA methods are provided in the sections that follow: listwise deletion, (re)weighting methods, pairwise deletion, the addition of indicator variables (or categories) to

[12] Note that SAS/STAT® 14.1 offers this functionality, but that version was not available at the time this research was conducted.

capture item nonresponse, and imputation. The literature indicates that pairwise deletion and the addition of indicator variables (or categories) are not recommended for use.

Further, the literature is mostly silent about when "missing values will not…be a serious source of concern" or when a simple imputation could be used in place of a missing value without a meaningful impact on the inference. An exception is Schafer (1999) who claimed that a missing rate of 5 percent or less was inconsequential. Bennett (2001) maintained that statistical analysis was likely to produce biased inferences (e.g., biased estimated regression coefficients) when more than 10 percent of values are missing.

In truth, the amount of missing item values should not be the sole criterion for assessing its impact on statistical analyses. As demonstrated convincingly by Groves and Peytcheva (2008), the missing item value mechanisms and the missing data patterns have greater impact on bias than the proportion of missing item values.

The literature often advises researchers to rely on their own understanding of the data to assess whether missingness is a cause for concern and whether treating imputed values as real undercuts the validity of the inference. One problem with this alternative is that a complex MIVRA method such as multiple imputation (MI; Section 3.2.5.1) or maximum likelihood (ML; Section 3.2.6) must be used to assess the impact on inference of not using the method.

There is also much in the literature about the need to do sensitivity analyses on the models underpinning model-based MIVRA methods like imputation (Section 3.2.5) and ML (Section 3.2.6), especially regarding the usual assumption that nonresponse is not a function of the missing item values (National Research Council, 2010). The same could be said of the response model used in weighting methods (Section 3.2.2).

Moreover, a monograph solely about missing item values in clinical trials (European Medicines Agency, 2010, p. 11) offers this useful advice about drawing inferences from a dataset with missing item values.

> [S]ensitivity analyses can be defined as a set of analyses where the missing data are handled in a different way as compared to the primary analysis. … When the results of the sensitivity analyses are consistent with the primary analysis and lead to reasonably similar estimates of the treatment effect, this provides some assurance that neither the lost information nor the methods used to handle missing data had an important effect on the overall study conclusions. In this situation, the robustness of the results is clear and the missing values will not…be a serious source of concern.

Finally, a distinction needs to be drawn between the analytic (regression) model being fit and the model used to account for missing item values. Both can fail. Many of the methods discussed in the remainder of Section 3.2, with a few notable exceptions, assume the analytic model being fit holds in the population of interest. Most assume the model used to account for the missing item itself is correct.

The available literature on specific MIVRA methods is summarized on the following pages.

3.2.1 Listwise Deletion

A popular method for handling item nonresponse in a statistical analysis that can account for a complex sampling design (i.e., the data can come from a stratified multistage sample and have unequal analysis weights) is listwise deletion. Also called "complete case analysis," this method simply removes all records with missing variables of interest from a particular analysis so that different analyses will use different datasets.

The problem with listwise deletion is that it discards incomplete records that may still contain useful information. This has led many to conclude, incorrectly, that listwise deletion is "among the worst methods available for practical applications" (Wilkinson & the Task Force on Statistical Inference, Board of Scientific Affairs, American Psychological Association, 1999, p. 598). In fact, the estimates after listwise deletion can sometimes be unbiased when those produced by its seemingly more sophisticated competitors are not (when item nonresponse in a covariate of the model, that is, a component of **x** in $\mathrm{E}(y) = \mathbf{x}^T \boldsymbol{\beta}$, is a function of the missing value itself; Kott, 2015).[13] Furthermore, many competing methods (i.e., MI and ML estimation) cannot as easily incorporate the impact of the sampling design.

3.2.2 Weighting Methods

It is often possible to adjust the analysis weights in a listwise-deleted dataset derived from complex survey data to remove the bias from the resulting estimates, even when nonresponse is a function of the dependent variable of the model. This is done by first estimating the probability (often called a "propensity") that a given record will be deleted in the listwise deletion process as a function (wholly or in part) of the value of the dependent variable. The inverse of that probability is then included as an additional factor in the weight for each record in the listwise-deleted dataset; that is, a record in the original dataset being retained in the listwise-deleted complex-survey dataset is treated as an additional phase of random sampling with a weighting factor estimated from the data.

A replication method can be used to measure the resulting standard error correctly. Ignoring the impact of this additional weighting step may theoretically bias the standard errors *upward* (making the statistical tests that use these standard errors conservative). This is because information used in the reweighting is not used in the standard error estimation.

A misconception is that item values must be missing completely at random (i.e., independent of all the variables in the model) for listwise deletion to produce unbiased parameter estimates in a regression analysis. This is only true when the model being fit does not hold in the population of interest. Even then, proper reweighting can greatly reduce the potential

---

[13] The terms "missing at random" and "not missing at random" (or "missing not at random") are mostly avoided because they are not very useful when fitting a regression model where records with missing item values are deleted. If the regression model holds in the population, and if the probability of a record being deleted is a function only of the covariates in the model and not of the dependent variable, then using listwise deletion will not lead to biased estimates even if item nonresponse is said to be not missing at random (i.e., dependent on the variables with missing values). On the other hand, if the probability of a record being deleted is a function of only the dependent variable, and that variable is never missing, then listwise deletion will lead to biased estimates even though nonresponse is said to be missing at random.

for bias. Moreover, most other MIVRA methods also produce biased estimates when the model being fit does not hold in the population.

The data summaries in Chapter 2 revealed that sometimes the NSDUH datasets that were analyzed included records with a missing value for the dependent variable. Some also included records whose membership in the subpopulation of interest was unknown. For those records, listwise deletion is often employed; that is, they are dropped from the subsequent analysis. Using listwise deletion on such records does not lead to biased results if the probability of a record being deleted is a function only of the independent variables of the model being fit and the analytic model fits the population of interest. Methods exist for handling the possibility that the missingness is a function of the dependent variable, most commonly as part of a sensitivity analysis, but they are beyond the scope of this report. More information on this topic is available in Section 5.4.

3.2.3 Pairwise Deletion

Pairwise deletion tries to retain some information that listwise deletion loses. When there are no missing item data, many estimates, such as those for linear regression coefficients, can be expressed as a function of estimated means, variances, and covariances. Pairwise deletion employs listwise deletion to compute each component of the estimate rather than the estimate itself (e.g., the numerator and denominator in a simple regression). Unfortunately, pairwise deletion cannot always be used, because either the estimate of interest cannot be expressed as a function of other estimates or the function cannot be carried out (e.g., a covariance matrix may not be invertible). Thus, pairwise deletion has few proponents and is not evaluated in the simulations discussed in Chapter 4. Allison (2002) provides more details on pairwise deletion.

3.2.4 Addition of an Indicator Variable (or Level of a Variable) to Denote a Missing Covariate

Another MIVRA method is adding an indicator variable to denote when a continuous variable has a missing value, or adding a category ("missing") when a categorical variable has a missing value (the two are mathematically equivalent). Jones (1996), however, demonstrated that this practice will lead to biased estimates of regression coefficients, even when the missingness is unrelated to the dependent variable, so long as a covariate (e.g., a component of **x** in the model $\mathrm{E}\left(y\right) = \mathbf{x}^T\boldsymbol{\beta}$ for the dependent variable *y*) with missing values is correlated with other covariates in the model, which is almost always the case. The size of the estimated coefficient of the covariate with missing values will tend to be underestimated, whereas the estimated coefficients of correlated covariates incorrectly try to "compensate" for the missing values.

All biases can be removed by putting additional covariates in the model to capture the possible interactions between the missingness indicator and the other covariates. Unfortunately, this technique would tend to undermine any efficiency gains from using this method in place of simple listwise deletion due to the increased number of parameters needing estimation.

Among the sources examined under this review, the authors were unable to identify documentation in the literature that indicated a covariate in a survey sample with a (weighted) fraction of missing values may in some sense accurately reflect the population (i.e., had the entire population been sampled, this fraction would be expected to respond "don't know" to a question about how easy is it to obtain a certain drug). In such a situation, a missing value may be a valid response. Therefore, it may be up to the analyst to decide whether "missing" is a valid response or whether it substitutes for a response that ideally should have been provided. Unless it is a valid response, it is not recommended that the response option of "don't know" is included as an indicator variable or a level of a variable in lieu of other methods intended to address item missingness.

3.2.5 Imputation

Imputation (i.e., replacing a missing value with a particular value) is perhaps the most common way to compensate for missing item values in a survey. Although not always explicitly stated, imputation makes assumptions about relationships between (or among) the variables in addition to those in the analytic model itself. The dependent variable in the analytic model can be—and often will be—a covariate in a probabilistic imputation model (i.e., the expected value under the imputation model of a component of **x** will be partly a function of y), as can the independent variables in the analytic model.

When the imputation model is probabilistic and correctly specified (i.e., $\mathrm{E}\left(x \mid \mathbf{z}\right) = \mathbf{z}^T\boldsymbol{\gamma}$, where x is a component of the covariates in the analytic model), imputing a missing item value by its expectation under the model will remove the bias from a resulting estimated mean or total of a particular variable. The same, however, cannot be said about an estimated regression model when covariate values are imputed. Adding appropriate random noise to the imputation will often fix that problem while not biasing estimated univariate means and totals.

Imputation with adjustment cells is a commonly used imputation method that separates the sample into mutually exclusive adjustment cells (for each missing item). A missing item value is imputed by randomly selecting a *donor* record from within the same cell from among the item respondents and then using the donor record's value in place of the missing value. (When a missing value is categorical, an equivalent approach uses the respondent values in the same cell to estimate the probabilities of falling into one of the categories and then randomly assigns the missing value using those probabilities.) Under a correctly specified imputation model in which every record in a cell is identically distributed, this donor imputation method can be used to remove the bias from both estimated means and estimated regression coefficients. This is the case, because the item value of a donor has the same expected value as the missing item value it replaces, and the difference between the donor's actual value and this common item mean is effectively a randomly selected error term.

Imputation models, such as the one in which every record in a cell is identically distributed, assume that the model holds equally well for respondents and nonrespondents of the item value in question. Moreover, the same model is usually assumed to hold for any record that would have been assigned to that cell, whether it is sampled or not. This last assumption can be relaxed by drawing the donor record with probability proportional to its sampling weight. This is what the weighted sequential hot-deck (WSHD) method does (Cox, 1980). Moreover, in addition to separating the sample into imputation cells, it sorts both the item respondents and nonrespondents by another variable and uses a complex selection routine not only to force the weighted distribution of the item values within a cell before and after imputation to be close (before: using only respondent values; after: using respondent and imputed values) but also to ensure that no donor is used more often than necessary for that near equality to hold.[14]

The two most common methods for separating the sample into imputation cells are "expert judgment," which can be informed by some testing, and recursive partitioning (also called "classification" or "regression trees"). Toth and Eltinge (2011) provide theoretical justification for the latter method with complex survey data.

3.2.5.1 Multiple Imputation (MI)

The standard errors of estimated regression coefficients cannot be estimated easily when each missing value is imputed with its expected value plus a random error. Treating imputed values as true values in standard error estimation almost always underestimates standard errors, even assuming the model used in the imputation is correct.

MI is one way to get around this problem while also increasing the accuracy of estimated regression coefficients. It does this by estimating each missing value several times (e.g., five times). Taking the average of the multiply imputed values removes much of the added variance due to the random noise added to each single imputation. In addition, computing the variability of the estimate across the five sets of imputed values provides the means for measuring the increase to standard error due to fitting models for the missing variables in the first place. Often in large-scale surveys that use MI procedures, multiply imputed datasets are released so that all analysts use the same datasets for their analyses, thus ensuring standardization across analyses.

Various articles in the literature (Kott, 1995; Kim, Brick, Fuller, & Kalton, 2006) have shown the limitations of MI with complex survey data. Appropriately incorporating the impacts of weighting, strata, and clustering into the imputation models is not always trivial. Even when the imputation successfully incorporates the impacts of the complex sampling design, the MI variance formula may still not provide asymptotically unbiased variance estimates. Nevertheless, except when estimating subpopulation means (which can be viewed as a form of regression analysis), computing MI estimates of standard error is often more reasonable (i.e., comes closer to being correct) than ignoring the impact of imputation on standard errors entirely.

3.2.5.2 Cycling (or Chaining)

One problem often faced when there are missing item values in a regression analysis is that the pattern of missingness across the independent variables can be quite complicated.[15] Such a situation makes it difficult to develop and fit a single multivariate model that successfully treats the dependent variable and all covariates with missing variables (as must be done with ML). The existence of complex nonresponse patterns is easily handled by listwise deletion and listwise deletion with reweighting. Unfortunately, the amount of data that is lost can be substantial and can result in estimates with large standard errors. In particular, even if no single variable is ever

---

[14] Recall that an ad hoc imputation method often used with NSDUH data creates imputation cells defined by age and race/ethnicity and then selects donors sorted by age with WSHD.

[15] With "n" independent variables, there are, in fact, $2^n - 1$ such missingness patterns that may possibly arise.

missing more than 2 percent of the time, up to 20 percent of the records in a regression analysis with 10 covariates can be incomplete.

Raghunathan, Lepkowski, Van Hoewyk, and Solenberger (2001) introduced a *cycling* (or *chaining*) procedure for MI that greatly simplified the problem. The idea of cycling is to develop imputation models one variable at a time, usually from the variable with the least number of missing values to the one with the most, using known but also previously imputed values for the other covariates. Once the process is completed, it begins again, dropping the imputed values only for the variable whose imputation model is being fit. Despite theoretical questions about whether this process converges, or even whether the univariate models are logically consistent, the method appears to work well in practice (van Buuren, Brand, Groothuis-Oudshoorn, & Rubin, 2006).

Although developed for MI, chaining can also be used with a single imputation for each missing value. Moreover, standard errors accounting for the modeling involved in these imputations can often be computed using a version of the Rao-Shao jackknife (Rao & Shao, 1992; Cohen, 2002), which requires not only replicate weights but also replicate imputations.

3.2.6 Maximum-Likelihood Estimation

An alternative to MI and to imputation in general is the direct use of ML estimation (also called "full-information ML," "case-based ML," or "casewise ML"). When fitting a model, this method requires that all variables with missing values in a model be treated as random variables in a larger and correctly specified model. Its greatest strength is that it can, in principle, use all available information efficiently (i.e., ML results in estimates with smaller standard errors).

Both MI and ML require specification not only of the models for the expected values of the dependent variable and covariates with missing values but also of their variance/covariance structure and often their probability distributions. Therefore, the resulting estimates and their standard errors are not as robust to model misspecification as is usually the case when analyzing survey data, where, for example, the errors in a linear regression model are rarely assumed to be normal and their variances need not be specified up to a constant (Skinner, 1989). That robustness can often be regained by treating the score function (the derivative of the log likelihood with respect to the model parameters, which equals zero when the likelihood is maximized) as an estimating equation and computing robust standard errors for its solution, most easily by replication. Incorporating the weights is trivial in this context because there is no longer a claim that the solution is ML, whereas clustering and stratification are treated as nuisances in robust variance estimation. The method is more properly called "pseudo-maximum likelihood" (Gourieroux, Monfort, & Trognon, 1984).

Using ML methods requires the analyst to specify a model structure (which can occur indirectly, in that the software may implicitly assume a model structure depending on the software employed). The same can be said about MI, except that the specification is usually the responsibility of those releasing the datasets, which may not agree with how individual analysts might approach it (e.g., the analyst may want to study the relationship between two variables that the data releasers had assumed were unrelated). Indeed, the same multiply imputed datasets supplied by the data owners can be used regardless of the analysis, whereas ML is analysis specific, as is listwise deletion.

3.3 Selection of MIVRA Methods for the Simulation Experiment

In the literature review described in Section 3.2, several MIVRA methods were examined. The methods deemed to be more promising were relatively straightforward to implement using readily available statistical software packages, could account for data from a complex survey, and could generate results that could be replicated by others. In this section, the decisions on which methods were deemed worthy of further investigation are described and justified.[16] The selected methods were evaluated empirically in a simulation study (Chapter 4) and included the following:

- *Listwise deletion with and without reweighting.* These are two straightforward methods that can be implemented using any statistical software package.
- *Hot-deck imputation.* Imputation methods are well known to many analysts, although they vary with respect to ease of implementation and replicability.
- *Maximum-likelihood estimation*. Although these methods are not as frequently used by many analysts, the literature review suggested that they are not too difficult to implement, are often replicable, and (especially) are effective for these types of situations.

Details about the implementation of these methods, including discussions of software options, are provided in Sections 3.3.1, 3.3.2, and 3.3.3, respectively. Section 3.3.4 lists the final six methods used in the simulation study.

3.3.1 Listwise Deletion with and without Reweighting

Listwise deletion (Section 3.2.1) meets the three criteria mentioned above, in that it is straightforward to implement using readily available statistical software packages, it can account for data from a complex survey, and it can generate results that are replicable by others.

Listwise deletion with reweighting (Section 3.2.2) also meets these three criteria, although it requires the extra step of adjusting the weights. The reweighting can be done in a variety of ways. In the simulation study, the WTADJUST procedure in SUDAAN® was used, though dichotomous logistic regression may also be used. The latter is a method frequently used by analysts of NSDUH data at present and can be implemented in all statistical software packages. All never-missing variables in the analytic model (including the dependent variable in the analytic model) were used as dependent variables in the calibration model. Sample code in SAS®, SAS-callable SUDAAN, and Stata® is provided in Appendix C.[17]

3.3.2 Hot-Deck Imputation

Two hot-deck imputation methods were evaluated in the simulation study: simple WSHD and cyclical tree-based hot deck (CTBHD). Although both methods as implemented in the

---

[16] MI was not evaluated in the simulation study. Due to its complexity, MI may be best performed by the statisticians creating the dataset for analysis. Moreover, due to the stochastic nature of MI, two analysts could get different results even when using the same imputation models, which makes reproducibility more difficult.

[17] In Appendix C, the SAS-callable SUDAAN sample code uses the WTADJUST procedure, and the SAS and Stata sample codes use dichotomous logistic regression. It is worth noting that R offers a comparable procedure to SUDAAN's WTADJUST procedure.

simulation study use SUDAAN, other statistical software packages offer these imputation routines as well. The WSHD method implemented in the simulation study is a simpler and more straightforward imputation method than the CTBHD method.

Simple WSHD, described in more detail in Appendix C, involves three steps: the formation of imputation cells, the sorting of item respondents and nonrespondents within imputation cells, and the assignment of a donor to each item nonrespondent. In this study, the imputation cells were somewhat arbitrarily defined by the cross-classification of gender and race/ethnicity (Hispanic/Latino, non-Hispanic/Latino white, non-Hispanic/Latino black/African American, and non-Hispanic/Latino other), and the sorting variable within each cell was age.[18] The complex donor selection routine associated with this approach has several nice properties described in Section 3.2.5.

The WSHD imputation method can be modified to use more data-driven processes to define donor cells and can also include cycling. That is what the CTBHD method evaluated here does. CTBHD selects distinct cells using regression trees (Section 3.2.5) and cycles two times through the variables with missing item values to capture more information from item respondents given a complex missingness pattern. Again, WSHD would be used to select donors within cells. This method (as well as the software used to implement it) is an RTI-developed donor imputation system originally designed for education surveys (Wine, Bryan, & Siegel, 2014). All variables in the analytic model were used to support the imputations for all the sometimes-missing model variables.

3.3.3 Maximum-Likelihood Estimation

ML estimation (Section 3.2.6) appears to have many virtues based on the literature review, but commonly used statistical software packages like SAS and Stata did not fully support ML for complex survey data at the time the analysis described in this report was undertaken.[19] The ideal software would offer the ability to perform ML on categorical or continuous data, with some missing values, from complex survey samples that include unequal weights, clustering, and stratification. The methods offered by SAS and Stata appear unsuitable for use in NSDUH for handling missing data via ML for the following reasons:

- The CALIS procedure in SAS supports categorical variables and performs ML but does not account for a complex survey design.
- The SEM command in Stata can account for a complex survey design and perform ML on missing values, but it supports only continuous variables, whereas most of the NSDUH variables used in analyses are categorical.
- The GSEM command in Stata supports categorical variables, but it does not support complex survey designs, which is critical to the NSDUH data, nor does it perform ML on missing values.
- The LOGIT and MLOGIT commands in Stata support categorical variables and can account for a complex survey design but do not perform ML on missing values.

[18] This is a method frequently used in previous regression analyses of NSDUH data. See Chapter 2.

[19] R has a package called lavaan.survey that has the potential to support ML for complex survey data with missing values, but no literature was found that mentions this functionality.

- An independently developed Stata command (GLLAMM) can support logistic regression but not in the context of a complex survey design. In particular, it (1) uses syntax that is quite different from Stata syntax, (2) runs very slowly (one of the models used in the simulation study took approximately 6 hours to run), and (3) has no user support other than discussion boards with other users.

Two other software packages, Mplus® and Latent GOLD®, were found to be capable of performing ML on categorical or continuous data from complex survey samples that include unequal weights, clustering, and stratification. Both Mplus and Latent GOLD were designed primarily for estimating structural-equation models with latent variables, but they also have the capability to estimate single-equation logistic models without latent variables (unless one considers a missing value latent) in the complex sample paradigm, which is why they were investigated in the simulation experiment. The "MPLUS" method is described in more detail in Appendix D, and the "LG" (Latent GOLD) method is described in more detail in Appendix E.

Table 3.1 summarizes the ML capabilities of SAS, Stata, Mplus, and Latent GOLD.

Table 3.1 Maximum-Likelihood Capabilities of Various Software Packages for NSDUH Regression Models

Software Package	Categorical Variables	Complex Survey Design	Maximum Likelihood for Missing Values
Mplus®	X	X	X
Latent GOLD®	X	X	X
SAS PROC CALIS	X		X
Stata® Commands			
SEM		X	X
GSEM	X		
LOGIT	X	X	
MLOGIT	X	X	
GLLAMM	X		X

3.3.4 MIVRA Methods Used in the Simulation Study

Based on the literature review and the software considerations described in Sections 3.3.1 to 3.3.3, the following six methods were chosen to investigate further in the simulation study as described in Chapter 4.

1. listwise deletion (LD)
2. listwise deletion plus reweighting, using SAS and SUDAAN (REWT)
3. weighted sequential hot deck using gender, race/ethnicity, and age as auxiliary variables (WSHD)
4. cyclical tree-based hot deck using all model variables to support the imputations (CTBHD)
5. ML using Mplus software (MPLUS)
6. ML using Latent GOLD software (LG)

4. Simulation Experiment, Results, and Interpretation

4.1 Introduction

This chapter compares the effectiveness of alternative methods for handling missing item values in regression analyses (MIVRA) using simulations based on National Survey on Drug Use and Health (NSDUH) data. The six models and the (sub)samples used in the simulation experiment, previously introduced in Table 2.2 of Section 2.3, are summarized in Table 4.1.[20] Table 4.1 also displays the number of variables in each of the six models investigated that did not have any item missingness (Never Missing) and the number of variables that did have item missingness (Sometimes Missing). Each of the models had between 26 and 35 estimated coefficients, not counting intercepts. Note that a single categorical variable can produce a number of coefficients, depending on the number of levels of each variable.

Table 4.1 Six Regression Models Used in Simulation Experiment

Study/Model	Dependent Variable (Dichotomous, unless Otherwise Stated)	NSDUH Subpopulation of Interest	Survey Years	Number of Variables (including Dependent Variable)		Approximate Subsample Size (Number of Records Having a Dependent Variable Value)
				Never Missing	Sometimes Missing	
N4/SPDMON	Past month SPD	Women aged 18-44	2008-2012	10	6	92,700
N4/MHTRT	Past year mental health treatment	Women aged 18-44, with past month SPD	2008-2012	10	8	7,600
N14/YOTMTHLP	How much did counseling help (on a scale from 1 to 5)?	12 to 17 year olds with past year MDE who received counseling in the past year	2006-2010	7	8	3,300
N14/YORXHLP	How much did mental health medications help (on a scale from 1 to 5)?	12 to 17 year olds with past year MDE who took medication in the past year	2006-2010	7	7	1,500

[20] The only differences between Table 4.1 and Table 2.2 are (1) the names given to the models, which will be used for the rest of this report; (2) the additional columns with the number of variables; and (3) the sample sizes in the rightmost column. The sample sizes in Table 2.2 include all records known to be in the subpopulation, whereas the sample sizes in Table 4.1 include only those records known to be in the subpopulation with a nonmissing value for the dependent variable. For the simulation study, all records with a missing value for the dependent variables were dropped from the dataset.

Table 4.1 Six Regression Models Used in Simulation Experiment (continued)

Study/Model	Dependent Variable (Dichotomous, unless Otherwise Stated)	NSDUH Subpopulation of Interest	Survey Years	Number of Variables (including Dependent Variable)		Approximate Subsample Size (Number of Records Having a Dependent Variable Value)
				Never Missing	Sometimes Missing	
N19/ANLYR	Past year use of pain relievers	12 to 17 year olds	2008-2012	6	7	112,600
N19/ABODANL	Past year pain reliever disorder	12 to 17 year olds who used pain relievers in the past year	2008-2012	6	7	6,300

MDE = major depressive episode; SPD = serious psychological distress.

These simulations were not produced by modeling the relationships among the variables. Instead, they were generated to preserve, as much as possible, the actual relationships in the NSDUH data. Because the NSDUH datasets available had missing item values, however, the first step was to create a clone dataset without item missingness for each of the six NSDUH datasets used for each of the six models (Section 4.2 provides details). By design, the clone dataset mimicked the NSDUH dataset it approximated by having the same size and a similar empirical covariance structure but was without missing item values. From the clone dataset, datasets with missing values similar to the actual NSDUH dataset could be generated.

The second step was to simulate item missingness in each clone dataset (i.e., the simulated dataset without missing values) in a way that approximated the actual missingness pattern in the corresponding real sample (Section 4.3). This was done up to 1,600 times for each of three deletion rates (as defined in Chapter 2): 5 percent, 12.5 percent, and 20 percent.

The final step, described in Section 4.4, was to apply MIVRA methods (e.g., listwise deletion) to each of the simulated datasets with missing values. Then, estimates from the simulated datasets created in step 1 (without item missingness) were compared with estimates from the datasets in step 2 (with item missingness) that were calculated using different MIVRA methods. For each MIVRA method, the parameter estimates (coefficient estimates and their standard errors) that were calculated using the clone dataset without item missingness were compared with the estimates calculated using the simulated datasets with item missingness.

Response modeling (i.e., modeling if there is a response on an item as opposed to the value of responses) was used in creating the clone datasets and simulating patterns of missingness. This method may appear to favor listwise deletion with reweighting, which also uses a response-modeling technique. However, that is not the case for two reasons. First, a multinomial logistic model was used in estimating and then simulating the occurrences of patterns of item missingness (variables 1, 3, and 5 being the only missing variables is an example of a missingness pattern). Reweighting, by contrast, is based on a simple binary logistic model, in that a record is either deleted because at least one item is missing or it is not. Second, the multinomial logistic model used assumed that the probability of a missingness pattern occurring was a function of all the variables in the model that were never missing, including the dependent variable. One advantage of listwise deletion, in general, over its competitors is that both the versions with and without reweighting allow item missingness to be a function of the missing item's value. That was assumed *not* to be the case in the simulations. The inclusion of the dependent variable among the variables in the multinomial logistic missingness model would tend to make the results from listwise deletion (without reweighting) biased. The magnitude of the bias would depend on how much the value of the dependent variable influenced the probabilities of item missingness.

The remainder of this chapter describes in detail how the processes of creating a clone dataset (Section 4.2.1) and inducing item missingness (Section 4.2.2) were executed, and then focuses on the candidate MIVRA methods and the actual measures of their effectiveness in Section 4.4.

4.2 Technical Details on the Implementation of the Simulation Experiment

This section provides some of the technical details of the simulation experiment. In particular, Section 4.2.1 discusses how a clone dataset with no missing values was created from each original analytic dataset. Section 4.2.2 describes the stochastic process by which nonmissing values in the clone dataset were replaced by missing values for each iteration of the simulation. Finally, as a lead-in to the discussion of results in Sections 4.3 and 4.4, Section 4.2.3 provides technical details on the measures used to evaluate the MIVRA methods. Effective methods produce regression coefficient estimates with low bias, low variance, and accurate variance estimates, but the methods by which these three measures are produced are not always straightforward.

4.2.1 Creating a Clone Sample

Under ideal circumstances (such as the absence of item nonresponse), one of the six regression models described in Table 4.1 would be estimated using a sample of records (k denotes a record) labeled S. Estimates would be derived for the coefficients of either a binary logistic regression model (i.e., the N4 and N19 models),

$$E(y_k \mid \mathbf{x}_k) = \frac{1}{1+\exp\left[-(\beta_0 + \mathbf{x}_{1k}{}^T \boldsymbol{\beta}_1 + \mathbf{x}_{2k}{}^T \boldsymbol{\beta}_2)\right]}, \qquad 4.1$$

or the proportional-odds regression model with $L > 2$ levels (for the N14 models, $L = 5$),

$$E(y_{\ell k} \mid \mathbf{x}_k) = \frac{1}{1+\exp\left[-(\beta_\ell + \mathbf{x}_{1k}{}^T \boldsymbol{\beta}_1 + \mathbf{x}_{2k}{}^T \boldsymbol{\beta}_2)\right]}, \quad \ell = 1, \ldots, L-1, \qquad 4.2$$

in the population from which S was drawn, where y_k and $y_{\ell k}$ are 0/1 dependent variables that are never missing. By construction in a proportional-odds regression, $y_{\ell k} \geq y_{(\ell-1)k}$ for $\ell > 1$; that is, the categories are cumulative (which is why this is sometimes called "a cumulative logistic regression").

For convenience, the explanatory vector of variables in both equations has been divided into two subvectors: $\mathbf{x}_k{}^T = \left(\mathbf{x}_{1k}{}^T \mathbf{x}_{2k}{}^T\right)$. ($T$ denotes the transpose of a vector or matrix.) None of the component variables of $\mathbf{x}_{1k}$ is ever missing, whereas each of the component variables of $\mathbf{x}_{2k}$ is missing from at least one record in S.

Preventing the immediate estimation of the model in either equation are the missing item values in the associated sample *S*. That is why a missing item data method is needed.

The sample analogue of *S* (i.e., the idealized sample with no missing item values), called *F*, is not known. Consequently, *F*-clone, called *F**, is created. Ideally, *F** would have roughly the same size as *F*, whereas the distributions of all variables in the model being estimated (i.e., equation 4.1 or 4.2) would be close to the same for both *F* and *F**.

The largest subset of *S* with no missing item values (among the components of $\mathbf{x}_{2k}$) is called *D*. The probability that an observation in *S* was also in *D* was estimated by assuming that the probability of an observation's inclusion in *D* had the logistic form

$$\Pr(k \in D \mid k \in S) = \frac{1}{1 + \exp\left[-(\gamma_0 + \mathbf{z}_k{}^T \gamma_1)\right]}, \qquad 4.3$$

where $\mathbf{z}_k{}^T = \left(\mathbf{y}_k{}^T, \mathbf{x}_{1k}{}^T\right)$ and $\mathbf{y}_k{}^T$ is y_k when equation 4.1 needed to be estimated and is $(y_{1k}\ y_{2k} \cdots y_{(\ell-1)k})$ when equation 4.2 needed to be estimated.

Assuming the *inclusion model* defined by equation 4.3 was true, the binary logistic regression model was fit using the sampling weights and the LOGISTIC procedure in SUDAAN®. This produced an adjustment factor greater than 1 for each observation in *D*. An observation's adjustment factor was the inverse of the SUDAAN-computed estimated probability that it was included in *D* based on its characteristics (i.e., the component of $\mathbf{z}_k$).

The dataset *F** was then created by drawing a sequential probability proportional to size (PPS) sample from *D* (using the PPS_SEQ option of SURVEYSELECT in SAS®), with the adjustment factors as the measure of size and the sampling interval equal to 1. This meant that many observations in *D* were repeated in *F** to compensate for the observations in *F* that had missing item values. The less likely an observation was included in *D* given its characteristics, the greater its adjustment factor and consequently its likeliness to be repeated in *F**.

The number of observations in *F** was exactly equal to the number of observations in *F*. The relationship among the *y*- and *x*-variables in *F** roughly mimics those in *F*. The mimicry would be asymptotically perfect if the logistic inclusion model for *D* in equation 4.3 were true. Even if untrue, the resulting *F** remained a reasonable sample to study because the distribution of model variables in *F** was reasonably close to the distributions in the real complete NSDUH sample *F*, certainly closer than using *D* as a proxy for *F*.

4.2.2 Estimating and Inducing Missingness

Nonresponse in *F** was simulated 1,600 times (see more details in Table 4.9). For each of the six regression models in Table 4.1, the same 1,600 simulated datasets were used for each of the missingness methods discussed. Each simulation was generated from an item missingness model that could have produced the missingness pattern observed in *S*. This was done by first renumerating every pattern, $p_g, g = 1, \ldots, G$, realized in *S* (e.g., only the first component of $\mathbf{x}_{2k}$ was missing, only the second and fourth components were missing, etc.), and then sorting the patterns by their frequency: most common (no items missing) to patterns that happened only once. A sample-weighted multinomial logistic model was then fit to include as many patterns as MULTILOG in SUDAAN would accept—again, starting with the most common pattern—and putting all the least common patterns into the same final "pattern," $p_{\bar{G}}$ (where $\bar{G} \le G$). The *missingness model* fit was defined by

$$p_k^g = \frac{\exp(\gamma_{0g} + \mathbf{z}_k{}^T \boldsymbol{\delta}_{1g})}{\sum_{h=1}^{G'} \exp(\gamma_{0h} + \mathbf{z}_k{}^T \boldsymbol{\delta}_{1h})}, \qquad 4.4$$

where p_k^g was the probability that observation *k* had missingness pattern *g*.

Table 4.2 displays the patterns used in fitting equation 4.4. As shown in Table 4.2, not all the variables used in creating the components of $\mathbf{z}_k$ in equation 4.3 were included in every model (note that a categorical variable with *m* levels contributed *m* – 1 components to $\mathbf{z}_k$). This was necessary to achieve model convergence. Tables 4.3 through 4.8 show how many distinct patterns of missingness appeared in equation 4.4. The bottom row in each table is for the final pattern that combined many rare patterns.

By fitting equation 4.4 to *S*, the probability of each missingness pattern was estimated given $\mathbf{z}_k$ except for the rare patterns in $p_{\bar{G}}$. The overall relative frequency of a pattern in $p_{\bar{G}}$ was used to allocate its estimated probability given $\mathbf{z}_k$ among the other rare patterns in $p_{\bar{G}}$.

After estimating missingness probabilities as described above, the estimated probabilities of all patterns with at least one missing value were scaled so that the expected deletion rate would be 5 percent (e.g., $\sum_{k \in F} \sum_{g=2}^{G} p_k^g / n = .05,$ where *n* is the size of *S*). The pattern with no missing values (p_k^1) was modified accordingly (i.e., so that $\sum_{g=1}^{G} p_k^g = 1$ for all *k*.)

This created one set of missingness probability patterns for the simulations. Two other sets of missingness probability patterns were created by scaling the estimated probabilities of patterns with at least one missing value so that the expected deletion rate would be 12.5 percent and 20 percent, respectively.

Depending on the set of missingness probability patterns, samples from *F** with missing values were simulated in the following manner. For each observation $k \in F^*$, a random number *u* was independently selected from the uniform distribution on (0,1). T_r was the sum of the observation's missingness probabilities for the first *r* patterns. If $T_{q-1} \le u < T_q$, with T_0 defined to be 0, then the q^{th} item nonresponse pattern was selected for observation *k**.

Finally, as noted in the introduction to this chapter, having item nonresponse in equation 4.4 be a function of the components of y_k and $\mathbf{x}_{1k}$, but not $\mathbf{x}_{2k}$, puts both listwise deletion and reweighted listwise deletion at a disadvantage when compared with other missing data methods. Recall that listwise deletion is not sensitive to item nonresponse that depends on the values of *any* of the components of $\mathbf{x}_k$, whereas reweighted listwise deletion need not be sensitive to item nonresponse that depends on the values of $\mathbf{x}_k$ and y_k.

The simple reason that components of $\mathbf{x}_{2k}$ were not included in the missingness model was that estimating nonresponse as a function of those components could not be done easily.[21] Moreover, even if equation 4.4 described the missingness pattern in *S*, an estimation of that pattern's parameters provides only estimates, not true parameter values. Despite these limitations, the simulations of the missingness patterns were useful proxies of real situations.

Table 4.2 Summary of Missingness Patterns in Analyzed Samples

Study/Model[1]	Sample Size	Overall Deletion Rate (%)	Number of Patterns in Model[2]	Number of Variables Retained in Response Model ($\mathbf{z}_k$ in Equation 4.4)
N4/SPDMON	92,578	4.37	5 (+1)	Kept 5 of 10
N4/MHTRT	7,583	4.71	3 (+1)	Kept 5 of 10
N14/YOTMTHLP	3,271	15.84	4 (+1)	Kept all 7
N14/YORXHLP	1,539	17.15	4 (+1)	Kept all 7
N19/ANLYR	112,591	13.14	29 (+1)	Kept all 6
N19/ABODANL	6,296	12.58	8 (+1)	Kept all 6

[1] See Table 4.1 for study and model details.
[2] The pattern "everything else" is denoted by (+1).

Table 4.3 Specific Missingness Patterns for N4/SPDMON

Pattern	Components of Subvector ($\mathbf{x}_2$) with Missing Item Values						Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*88,644*	*95.63*
2	*X*	*.*	*X*	*X*	*X*	*X*	*2,503*	*2.70*
3	*X*	*X*	*X*	*.*	*X*	*X*	*774*	*0.83*
4	*X*	*X*	*X*	*X*	*X*	*.*	*370*	*0.40*
5	*X*	*X*	*X*	*X*	*.*	*X*	*287*	*0.31*
6	.	X	X	X	X	X	24	0.03
7	X	X	X	X	.	.	16	0.02
8	X	.	X	X	X	.	15	0.02
9	X	X	X	.	X	.	14	0.02
10	X	X	.	X	X	X	13	0.01
11	X	.	X	.	X	X	13	0.01
12	X	X	X	.	.	X	12	0.01

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-5) in the table denote the missingness patterns that were modeled. Not shown are four additional missingness patterns with single-digit frequencies.
Note: See Table 4.1 for N4/SPDMON details.

[21] A possible method is discussed briefly in Chapter 5.

Table 4.4 Specific Missingness Patterns for N4/MHTRT

Pattern	Components of Subvector (x_2) with Missing Item Values								Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*7,226*	*95.29*
2	*X*	*.*	*X*	*X*	*X*	*X*	*X*	*X*	*175*	*2.31*
3	*X*	*X*	*X*	*X*	*X*	*.*	*X*	*X*	*63*	*0.83*
4	X	X	X	.	X	X	.	.	57	0.75
5	X	X	X	X	.	X	X	X	33	0.44
6	X	X	X	.	X	X	X	X	16	0.21

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-3) in the table denote the missingness patterns that were modeled. Not shown are six additional missingness patterns with single-digit frequencies.

Note: See Table 4.1 for N4/MHTRT details.

Table 4.5 Specific Missingness Patterns for N14/YOTMTHLP

Pattern	Components of Subvector (x_2) with Missing Item Values								Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*2,753*	*84.16*
2	*X*	*X*	*X*	*X*	*X*	*X*	*.*	*X*	*199*	*6.08*
3	*X*	*.*	*X*	*X*	*X*	*X*	*X*	*X*	*176*	*5.38*
4	*X*	*X*	*X*	*X*	*X*	*.*	*X*	*X*	*67*	*2.05*
5	X	.	X	X	X	X	.	X	23	0.70
6	X	X	X	.	X	X	X	X	11	0.34

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-4) in the table denote the missingness patterns that were modeled. Not shown are 18 additional missingness patterns with single-digit frequencies.

Note: See Table 4.1 for N14/YOTMTHLP details.

Table 4.6 Specific Missingness Patterns for N14/YORXHLP

Pattern	Components of Subvector (x_2) with Missing Item Values							Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*1,275*	*82.85*
2	*X*	*X*	*X*	*X*	*X*	*X*	*.*	*121*	*7.86*
3	*X*	*.*	*X*	*X*	*X*	*X*	*X*	*77*	*5.00*
4	*X*	*X*	*X*	*X*	*X*	*.*	*X*	*31*	*2.01*

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-4) in the table denote the missingness patterns that were modeled. Not shown are 13 additional missingness patterns with single-digit frequencies.

Note: See Table 4.1 for N14/YORXHLP details.

Table 4.7 Specific Missingness Patterns for N19/ANLYR

Pattern	Components of Subvector (x_2) with Missing Item Values							Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*97,802*	*86.86*
2	*X*	*.*	*X*	*X*	*X*	*.*	*X*	*5,689*	*5.05*
3	*X*	*X*	*X*	*X*	*X*	*X*	*.*	*1,256*	*1.12*
4	*X*	*X*	*X*	*X*	*X*	*.*	*X*	*1,173*	*1.04*
5	*.*	*X*	*X*	*X*	*X*	*X*	*X*	*1,098*	*0.98*
6	*X*	*.*	*X*	*X*	*X*	*X*	*X*	*1,017*	*0.90*
7	*X*	*X*	*X*	*.*	*X*	*X*	*X*	*1,009*	*0.90*

Table 4.7 Specific Missingness Patterns for N19/ANLYR (continued)

Pattern	Components of Subvector ($\mathbf{x}_2$) with Missing Item Values							Number	Percent
8	*X*	*X*	.	*X*	*X*	*X*	*X*	*733*	*0.65*
9	*X*	.	*X*	*X*	*X*	.	.	*312*	*0.28*
10	*X*	*X*	*X*	*X*	.	*X*	*X*	*271*	*0.24*
11	.	*X*	*X*	*X*	*X*	*X*	.	*259*	*0.23*
12	.	.	*X*	*X*	*X*	.	*X*	*185*	*0.16*
13	.	.	*X*	*X*	*X*	.	.	*140*	*0.12*
14	.	.	*X*	*X*	.	.	.	*113*	*0.10*
15	*X*	*X*	*X*	*X*	*X*	.	.	*112*	*0.10*
16	*X*	.	*X*	*X*	*X*	*X*	.	*105*	*0.09*
17	.	*X*	*X*	*X*	*X*	.	*X*	*94*	*0.08*
18	.	.	*X*	*X*	*X*	*X*	*X*	*80*	*0.07*
19	*X*	.	*X*	*X*	.	.	*X*	*77*	*0.07*
20	*X*	.	*X*	.	*X*	.	*X*	*71*	*0.06*
21	*X*	.	*X*	*X*	.	.	.	*64*	*0.06*
22	*X*	.	.	*X*	*X*	.	*X*	*64*	*0.06*
23	*X*	*X*	.	.	*X*	*X*	*X*	*59*	*0.05*
24	*X*	*X*	*X*	*X*	.	*X*	.	*54*	*0.05*
25	.	*X*	*X*	*X*	.	*X*	*X*	*54*	*0.05*
26	*X*	*X*	*X*	*X*	.	.	*X*	*48*	*0.04*
27	.	.	*X*	*X*	*X*	*X*	.	*39*	*0.03*
28	.	*X*	*X*	.	*X*	*X*	*X*	*37*	*0.03*
29	*X*	.	*X*	*X*	.	*X*	*X*	*35*	*0.03*
30	.	.	X	X	.	X	.	34	0.03
31	.	X	X	X	.	X	.	28	0.02
32	.	.	X	X	.	.	X	28	0.02
33	X	.	X	X	.	X	.	26	0.02
34	X	X	.	X	X	X	.	25	0.02
35	.	X	X	X	X	.	.	25	0.02
36	.	.	X	X	.	X	X	25	0.02
37	X	X	X	.	X	X	.	24	0.02
38	.	X	.	X	X	X	X	24	0.02
39	X	X	X	.	X	.	X	22	0.02
40	X	X	X	X	.	.	.	19	0.02
41	.	X	X	X	.	.	X	14	0.01
42	.	X	X	X	.	.	.	13	0.01
43	.	X	X	.	X	X	.	12	0.01
44	X	.	X	.	X	.	.	11	0.01

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-29) in the table denote the missingness patterns that were modeled. Not shown are 62 additional missingness patterns with single-digit frequencies.

Note: See Table 4.1 for N19/ANLYR details.

Table 4.8 Specific Missingness Patterns for N19/ABODANL

Pattern	Components of Subvector ($\mathbf{x}_2$) with Missing Item Values							Number	Percent
1	*X*	*X*	*X*	*X*	*X*	*X*	*X*	*5,504*	*87.42*
2	*X*	.	*X*	*X*	*X*	.	*X*	*256*	*4.07*
3	*X*	*X*	*X*	.	*X*	*X*	*X*	*143*	*2.27*
4	.	*X*	*X*	*X*	*X*	*X*	*X*	*67*	*1.06*
5	*X*	.	*X*	*X*	*X*	*X*	*X*	*55*	*0.87*
6	*X*	*X*	.	*X*	*X*	*X*	*X*	*49*	*0.78*
7	*X*	*X*	*X*	*X*	*X*	.	*X*	*46*	*0.73*
8	*X*	*X*	*X*	*X*	*X*	*X*	.	*44*	*0.70*
9	X	X	X	X	.	X	X	18	0.29
10	.	X	X	X	X	X	.	14	0.22

X = component is not missing; period symbol (.) = component is missing. (No component of $\mathbf{x}_{(1)}$ is ever missing.)

Note: The italicized rows (1-8) in the table denote the missingness patterns that were modeled. Not shown are 40 additional missingness patterns with single-digit frequencies.

Note: See Table 4.1 for N19/ABODANL details.

4.2.3 Evaluating MIVRA on Measures of Bias and Variability

Given a model, let c be a (nearly) unbiased estimated coefficient of the model computed on its F*, and let c_s be the estimate of the same coefficient computed using a particular MIVRA method on sample S_t ($t = 1, \ldots, T$). A measure of the *empirical bias* of the method in estimating the coefficient is

$$B = \sum^T (c_{ts} - c) / T .$$

A measure of the *relative empirical bias* of the method in estimating the coefficient is

$$relB = B / c .^{22} \qquad 4.5$$

Whether the method is actually biased can be determined by testing whether B is significantly different from 0. To do that, the sampling variance of the T values of c_s is computed first,

$$v_a = [\sum c_s^2 - (\sum c_s)^2 / T] / (T-1),$$

and then the t-statistics are computed,

$$t = B / \sqrt{(v_a / T)} \qquad 4.6$$

(because the estimated standard error of the random variable B is $\sqrt{(v_a / T)}$).

The value v_a (the "added variance") is also an estimate of the contribution of the method to the variance of c_s as an estimate for the coefficient. A measure for the *total variance* of c_s is

$$v = v_f + v_a , \qquad 4.7$$

where v_f is the estimated variance of c derived using the full-sample clone F*.

[22] As a percentage, the relative bias is *relB* × 100 percent.

A slightly biased estimate for the *total mean squared error* of c_s as an estimator for the coefficient is

$$m = v_f + v_a + B^2 = v_f + \sum(c_s - c)^2 / (T-1) . \tag{4.8}$$

This is slightly (but ignorably) biased because B^2 is a slightly biased estimate for the squared bias of c_s (because $E\left(B^2\right) = \left[E\left(B\right)\right]^2 + Var(B)$). Its estimated bias is the very small value v_a / T when *T* is large.

If this method were used in practice, one would not have *T* simulations; rather, there would be only one realization of an incomplete dataset. The variance estimate for c_s using this method (which may not estimate the variance or mean squared error well) is e_s. Its average value across all *T* simulations is

$$e = \sum e_s / T . \tag{4.9}$$

4.2.4 Summarizing the Number of Simulation Iterations for Each MIVRA Method

Table 4.9 summarizes the number of simulations used to evaluate each method under each model and subpopulation. It uses the term "deletion rates," even though a whole record is deleted only when LD or REWT is used. Moreover, because item missingness was generated randomly from a multinomial logistic model, these rates are, in fact, expected rates, not actual rates. As indicated in the table and noted in Section 4.2.2, as many as 1,600 simulated datasets were used for each of the missingness methods. The CTBHD and LG methods were added to the simulation study at a later date, so fewer simulations were used. Also, the LG method was implemented only for a deletion rate of 20 percent.

Table 4.9 Number of Simulation Iterations by MIVRA Method and Model

	Method (Number of Simulations Used in Experiment)					
Study/Model	**LD**	**REWT**	**WSHD**[1]	**MPLUS**[2]	**CTBHD**[3]	**LG**
N4/SPDMON[4]	1,600	1,600	1,600	1,600	400	49
N4/MHTRT[5]	1,600	1,600	1,600	1,600	400	None
N14/YOTMTHLP[6]	1,600	1,600	1,600	1,600	400	None
N14/YORXHLP[7]	1,600	1,600	1,600	1,600	400	None
N19/ANLYR[8]	1,600	1,600	1,600	None	400	49
N19/ABODANL[9]	1,600	1,600	1,600	1,600	400	None

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = Latent GOLD; MDE = major depressive episode; REWT = listwise deletion with reweighting; SPD = serious psychological distress; WSHD = weighted sequential hot deck.

[1] WSHD employs weighted sequential hot-deck donor imputation within cells determined by gender and race/ethnicity and sorted by age.

[2] MPLUS is a (pseudo)-maximum-likelihood MIVRA method in a standard program.

[3] CTBHD employs WSHD donor imputation within cells grown from two cycles of a regression tree.

[4] N4/SPDMON is a model of SPD in the past month among women aged 18 to 44.

[5] N4/MHTRT is a model of mental health treatment among women aged 18 to 44 with SPD in the past month.

[6] N14/YOTMTHLP is a model measuring how much counseling helped adolescents with an MDE in the past year who sought counseling in the past year.

[7] N14/YORXHLP is a model measuring how much medication helped adolescents with an MDE in the past year who used counseling in the past year.

[8] N19/ANLYR is a model of past year pain reliever use among adolescents.

[9] N19/ABODANL is a model of past year drug dependence among adolescents with past year pain reliever use.

4.2.5 Implementing the CTBHD Method

Table 4.10 displays the arbitrary stopping rules used to determine when to terminate further division of cells when employing the CTBHD methods, which forms classification and regression trees.

Table 4.10 Number of Variables (Including Dependent Variable)

Study/Model	Stopping Rule for CTBHD Cells: No Further Division when…	
	Cell Size Would Be Less than	Cell Size Is Already Less than
N4/SPDMON	100	200
N4/MHTRT	100	200
N14/YOTMTHLP	25	50
N14/YORXHLP	100	200
N19/ANLYR	100	200
N19/ABODANL	100	200

CTBHD = cyclical tree-based hot deck.
Note: See Table 4.1 for study and model details.

4.3 Summarizing the Simulation Results

Findings were summarized for each model and deletion rate by looking at the *distributions* across the coefficients of

- the relative biases and *t* values of the coefficient estimates;
- the ratios of the estimated total variance and mean squared error of each coefficient estimate to the estimated variance of the coefficient's estimate computed from the completed sample; and
- the relative bias of the variance estimate when using this method when viewed as an estimate for total variance and as an estimate for total mean squared error.

The summaries that appear in this chapter are for a deletion rate of 20 percent. Results for the other deletion rates are analogous but not as strong, as expected. They appear in Appendix F. Chapter 5 provides guidance on how to evaluate the extent of deletion/missingness for regression analyses.

4.3.1 Bias in Model Coefficient Estimates

Although Table 4.11 reports (weighted) medians, means, quartiles, minimums, and maximums of the empirical absolute relative biases for each model and method among the estimated coefficients of the analytic model (the sample size), the focus in this discussion is on the means and medians of these relative biases.

All the methods were expected to perform reasonably well (for bias) except, perhaps, LD and WSHD. LD does not use any auxiliary variables to correct for bias, and WSHD uses only age, gender, and race. REWT, by contrast, uses all the never-missing variables in the analytic model to correct for bias. The CTBHD, MPLUS, and LG methods use all the variables in the analytic model to correct for bias, including the ones with missing values.

Table 4.11 shows that MPLUS tends to perform well for the four models with small sample sizes, unlike the relatively poor performance of MPLUS for the two large models. For the model N4/SPDMON, it performed relatively poorly (i.e., had large empirical absolute relative biases), and N19/ANLYR, which would not run except on a subsample. It had the lowest median for three of the small models and the second-lowest median for the fourth. CTBHD and REWT also performed reasonably well. CTBHD never ranked lower than third and was the best method for two of the six analytic models; REWT was the best method for one of the models and the second-best for three of the others. LG performed poorly for N19/ANLYR; for N4/SPDMON, its moderate performance placed it in between the other methods. LD and WSHD performance tended to be worse than all the others, as expected. LD was the worst method for three of the six models, and WSHD was the worst for one and the second-worst for two of the others. These results largely met study expectations.

Table 4.11 summarizes the absolute values of the empirical relative biases ($|relB|$ in equation 4.5) of the coefficient estimates by model and method. By focusing on the absolute empirical relative biases of the estimated coefficients, the table summarizes how using a method tends to bias coefficient estimation. The summaries were computed by taking a weighted distribution across the coefficients of the model. The weights used were the absolute values of the estimated coefficients in the completed sample (i.e., $|c|$). These weights were used so that estimated coefficients with very small values did not dominate the summaries. The absolute size of the coefficient, as measured with an unbiased estimate, determined its contribution to the summary statistic.

For example, the top left part of Table 4.11 shows distributional statistics for the absolute empirical relative biases of the regression coefficients in the N4/SPDMON model, when LD was the MIVRA method and the deletion rate was 20 percent. There were 32 regression coefficients associated with the N4/SPDMON model. Using the full-sample clone F^* (Section 4.2.1), the point estimate of the regression coefficient associated with whether the woman had exactly one biological child in the household was very close to 0 (0.00138). Across the 1,600 iterations of the simulation, the mean of the point estimates for this regression coefficient was too low: -0.03444. Thus, following equation 4.5, the maximum absolute empirical relative bias associated with this regression coefficient was

$$\left|\frac{-0.03444-0.00138}{0.00138}\right| = 25.96754\,.$$

This corresponds to the value of 25.968 in Table 4.11 (25.968 × 100% = 2,596.8%), which is the coefficient with the largest absolute empirical relative bias among the 32 regression coefficients in the model. Note that weighting by $|c|$ (which is nearly 0) mitigates the impact of this value on the mean and the percentiles but has no impact on the maximum or minimum absolute empirical relative biases.

Table 4.11 Weighted Distribution of Absolute Empirical Relative Biases by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute Empirical Relative Bias					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	0.004	0.003	0.001	0.004	0.001	0.000
	1st Quartile	0.014	0.003	0.013	0.031	0.009	0.008
	Median	0.032	0.011	0.015	0.033	0.010	0.015
	Mean	0.059	0.048	0.022	0.092	0.019	0.021
	3rd Quartile	0.062	0.074	0.033	0.136	0.032	0.022
	Maximum	25.968	12.914	1.133	5.583	2.520	0.835
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	0.001	0.004	0.000	0.001	0.001	N/A
	1st Quartile	0.021	0.013	0.011	0.002	0.003	N/A
	Median	0.062	0.044	0.034	0.007	0.006	N/A
	Mean	0.123	0.079	0.050	0.014	0.022	N/A
	3rd Quartile	0.162	0.096	0.052	0.009	0.026	N/A
	Maximum	17.882	7.528	0.811	0.496	0.755	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	0.001	0.001	0.001	0.000	0.001	N/A
	1st Quartile	0.010	0.006	0.004	0.002	0.006	N/A
	Median	0.027	0.009	0.011	0.004	0.011	N/A
	Mean	0.040	0.023	0.036	0.015	0.037	N/A
	3rd Quartile	0.036	0.027	0.031	0.011	0.054	N/A
	Maximum	2.525	0.994	1.747	1.123	0.999	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	0.002	0.001	0.000	0.000	0.001	N/A
	1st Quartile	0.017	0.015	0.006	0.003	0.011	N/A
	Median	0.041	0.018	0.017	0.007	0.015	N/A
	Mean	0.075	0.035	0.034	0.018	0.030	N/A
	3rd Quartile	0.084	0.040	0.068	0.021	0.060	N/A
	Maximum	0.650	0.220	0.308	0.456	0.340	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	0.002	0.001	0.000	N/A	0.000	0.008
	1st Quartile	0.007	0.002	0.020	N/A	0.016	0.070
	Median	0.016	0.003	0.029	N/A	0.022	0.138
	Mean	0.045	0.025	0.044	N/A	0.039	0.288
	3rd Quartile	0.080	0.016	0.061	N/A	0.062	0.340
	Maximum	1.061	0.291	0.193	N/A	0.836	1.965
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	0.003	0.001	0.018	0.000	0.000	N/A
	1st Quartile	0.012	0.004	0.034	0.002	0.011	N/A
	Median	0.034	0.013	0.046	0.003	0.026	N/A
	Mean	0.071	0.033	0.066	0.020	0.055	N/A
	3rd Quartile	0.084	0.036	0.079	0.009	0.080	N/A
	Maximum	0.764	0.387	0.315	0.451	0.373	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across coefficients of $\left|relB\right|$ from equation 4.5, with the coefficients' $\left|c\right|$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., .032 is 3.2%).

Note: See Table 4.1 for study and model details.

Table 4.12 summarizes the t values of the empirical biases (equation 4.6) of the estimated coefficients by model and MIVRA method. The absolute values of the t values are summarized, and they were again weighted across all the coefficients in a model by the absolute values of the estimated coefficients in the completed sample ($|c|$).

A cause of some concern is that, for every method and model, the mean and median t values in Table 4.12 are greater than 2. That is, the bias due to item nonresponse is always significant (on average), if not necessarily large, no matter what method is used to try to remove it.

Table 4.12 Weighted Distribution of Absolute *T* Values of Empirical Biases by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute *T* Value of Empirical Bias					
		LD	**REWT**	**WSHD**	**MPLUS**	**CTBHD**	**LG**
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	2.476	2.763	1.481	6.363	0.993	0.032
	1st Quartile	12.949	3.609	23.492	57.433	10.658	2.577
	Median	22.750	10.878	90.894	194.826	25.787	12.210
	Mean	25.135	19.837	72.533	153.206	25.095	20.648
	3rd Quartile	33.291	22.265	116.023	227.041	41.821	35.271
	Maximum	68.406	76.152	159.515	345.048	61.504	69.634
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	0.368	0.307	0.023	0.444	0.157	N/A
	1st Quartile	15.266	5.568	9.356	3.542	1.744	N/A
	Median	21.367	13.160	19.313	8.685	3.879	N/A
	Mean	26.278	17.462	36.141	8.369	4.416	N/A
	3rd Quartile	31.277	31.411	87.820	10.005	6.514	N/A
	Maximum	63.584	55.091	91.392	38.331	16.159	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	0.173	0.234	1.192	0.030	0.415	N/A
	1st Quartile	2.393	1.074	5.496	1.544	2.859	N/A
	Median	5.588	2.247	12.665	5.410	5.223	N/A
	Mean	5.920	3.183	14.494	6.005	7.240	N/A
	3rd Quartile	7.518	3.723	18.359	9.006	9.718	N/A
	Maximum	19.713	12.294	66.749	18.234	31.727	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	0.372	0.263	0.004	0.133	0.387	N/A
	1st Quartile	2.371	2.004	4.899	3.441	2.811	N/A
	Median	7.878	3.833	10.609	5.189	5.014	N/A
	Mean	8.876	4.086	12.040	6.158	5.281	N/A
	3rd Quartile	12.527	6.116	18.939	8.628	7.697	N/A
	Maximum	22.179	9.348	26.086	21.111	13.430	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	0.686	0.229	0.491	N/A	0.461	1.312
	1st Quartile	6.546	1.197	24.079	N/A	8.767	3.783
	Median	10.165	2.703	63.228	N/A	19.070	15.927
	Mean	19.831	10.166	64.964	N/A	22.927	53.117
	3rd Quartile	23.152	4.887	95.353	N/A	31.840	114.318
	Maximum	85.989	64.777	220.559	N/A	76.552	165.231

Table 4.12 Weighted Distribution of Absolute *T* Values of Empirical Biases by MIVRA Method for 20 Percent Deletion Rate (continued)

Study/Model	Summary Statistics	Absolute *T* Value of Empirical Bias					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	0.517	0.181	5.554	0.000	0.217	N/A
	1st Quartile	5.180	1.024	15.803	2.202	3.735	N/A
	Median	11.082	3.220	23.305	3.051	7.397	N/A
	Mean	10.913	3.617	29.052	5.416	8.216	N/A
	3rd Quartile	18.239	5.328	31.974	4.380	10.404	N/A
	Maximum	27.099	12.607	73.738	21.913	26.993	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across the $|t|$ from equation 4.6, with the coefficients' $|c|$ values used as the weights.

Note: See Table 4.1 for study and model details.

4.3.2 Variance of Model Coefficient Estimates and Bias in Variance Estimates

Table 4.13 summarizes by model and MIVRA method the ratio of the empirical measure of total variance resulting from using the method (v in equation 4.7) to the estimated variance had there been no nonresponse (denoted by v_f). Table 4.14 replaces the empirical total-variance measure in the numerator with the empirical mean squared error (m in equation 4.8). The results in both tables are weighted across all the estimated coefficients in a model by the standard errors of the estimates had there been no response (i.e., $\sqrt{v_f}$).

The ratios in Tables 4.13 and 4.14 are always greater than 1. The difference between the values in those tables and unity measures the increase in variance and mean squared error, respectively, due to adjusting for item nonresponse with the method under investigation. The increase in standard error and root mean squared error is roughly half this amount because $(1+x)^{1/2} \approx 1 + x/2$ when x is between 0 and 0.4.

Table 4.13 Ratios of Empirical Variance to Full-Sample Variance by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical Variance to Full-Sample Variance					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	1.112	1.141	1.001	1.003	1.001	1.000
	1st Quartile	1.202	1.252	1.002	1.007	1.002	1.001
	Median	1.221	1.269	1.004	1.008	1.004	1.003
	Mean	1.217	1.266	1.020	1.038	1.021	1.010
	3rd Quartile	1.239	1.294	1.008	1.013	1.008	1.009
	Maximum	1.319	1.359	1.189	1.241	1.197	1.064

Table 4.13 Ratios of Empirical Variance to Full-Sample Variance by MIVRA Method for 20 Percent Deletion Rate (continued)

Study/Model	Summary Statistics	Ratio of Empirical Variance to Full-Sample Variance					
		LD	**REWT**	**WSHD**	**MPLUS**	**CTBHD**	**LG**
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	1.129	1.151	1.012	1.005	1.009	N/A
	1st Quartile	1.200	1.241	1.020	1.009	1.018	N/A
	Median	1.217	1.263	1.024	1.011	1.019	N/A
	Mean	1.221	1.259	1.043	1.021	1.037	N/A
	3rd Quartile	1.248	1.282	1.031	1.017	1.029	N/A
	Maximum	1.307	1.335	1.225	1.143	1.229	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	1.144	1.146	1.003	1.002	1.003	N/A
	1st Quartile	1.175	1.184	1.004	1.002	1.004	N/A
	Median	1.187	1.196	1.007	1.005	1.007	N/A
	Mean	1.192	1.199	1.039	1.024	1.039	N/A
	3rd Quartile	1.208	1.213	1.067	1.035	1.061	N/A
	Maximum	1.264	1.262	1.158	1.096	1.159	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	1.157	1.165	1.003	1.003	1.003	N/A
	1st Quartile	1.174	1.182	1.007	1.005	1.007	N/A
	Median	1.180	1.192	1.010	1.008	1.009	N/A
	Mean	1.201	1.215	1.040	1.028	1.044	N/A
	3rd Quartile	1.206	1.229	1.057	1.029	1.050	N/A
	Maximum	1.365	1.404	1.203	1.130	1.202	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	1.163	1.173	1.004	N/A	1.004	1.004
	1st Quartile	1.199	1.223	1.009	N/A	1.008	1.017
	Median	1.219	1.248	1.067	N/A	1.064	1.066
	Mean	1.229	1.256	1.076	N/A	1.076	136.220
	3rd Quartile	1.248	1.277	1.117	N/A	1.117	1.119
	Maximum	1.374	1.403	1.184	N/A	1.177	4030.381
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	1.159	1.172	1.006	1.003	1.006	N/A
	1st Quartile	1.171	1.185	1.011	1.006	1.011	N/A
	Median	1.209	1.232	1.056	1.035	1.052	N/A
	Mean	1.219	1.244	1.066	1.042	1.070	N/A
	3rd Quartile	1.244	1.260	1.091	1.054	1.098	N/A
	Maximum	1.354	1.382	1.183	1.134	1.202	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across coefficients of v / v_f from equation 4.7, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table 4.14 Ratios of Empirical MSE to Full-Sample Variance by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical MSE to Full-Sample Variance					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	1.198	1.195	1.002	1.006	1.001	1.001
	1st Quartile	1.254	1.277	1.004	1.009	1.003	1.003
	Median	1.306	1.312	1.013	1.031	1.006	1.009
	Mean	1.346	1.392	1.049	1.800	1.047	1.039
	3rd Quartile	1.364	1.397	1.072	1.411	1.041	1.069
	Maximum	1.979	2.321	1.408	7.123	1.380	1.224
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	1.190	1.233	1.013	1.006	1.010	N/A
	1st Quartile	1.249	1.254	1.023	1.009	1.018	N/A
	Median	1.278	1.279	1.028	1.012	1.020	N/A
	Mean	1.312	1.313	1.072	1.024	1.042	N/A
	3rd Quartile	1.307	1.331	1.069	1.017	1.039	N/A
	Maximum	2.015	1.882	1.487	1.198	1.266	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	1.148	1.146	1.003	1.002	1.003	N/A
	1st Quartile	1.179	1.185	1.004	1.003	1.004	N/A
	Median	1.191	1.197	1.010	1.005	1.010	N/A
	Mean	1.199	1.202	1.048	1.026	1.049	N/A
	3rd Quartile	1.213	1.215	1.071	1.035	1.064	N/A
	Maximum	1.285	1.286	1.243	1.115	1.254	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	1.161	1.167	1.003	1.003	1.003	N/A
	1st Quartile	1.179	1.184	1.008	1.005	1.008	N/A
	Median	1.196	1.195	1.010	1.008	1.011	N/A
	Mean	1.216	1.218	1.045	1.029	1.048	N/A
	3rd Quartile	1.223	1.235	1.065	1.029	1.054	N/A
	Maximum	1.397	1.420	1.237	1.131	1.220	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	1.163	1.182	1.005	N/A	1.004	1.006
	1st Quartile	1.205	1.224	1.067	N/A	1.062	1.072
	Median	1.239	1.259	1.153	N/A	1.125	1.175
	Mean	1.367	1.327	1.180	N/A	1.154	151.431
	3rd Quartile	1.326	1.324	1.196	N/A	1.182	1.873
	Maximum	3.103	2.254	1.701	N/A	1.701	4171.884
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	1.159	1.173	1.006	1.003	1.006	N/A
	1st Quartile	1.173	1.186	1.014	1.007	1.013	N/A
	Median	1.227	1.233	1.072	1.035	1.060	N/A
	Mean	1.245	1.250	1.087	1.046	1.083	N/A
	3rd Quartile	1.285	1.283	1.126	1.057	1.108	N/A
	Maximum	1.417	1.400	1.286	1.175	1.237	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across coefficients of m / v_f from equation 4.8, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table 4.14 quantifies the amount of mean squared error that is introduced by the bias. For example, when estimating the model N4/MHTRT using the WSHD method with a 20 percent deletion rate, the average (empirical) total mean squared error is roughly 2.9 percent higher than the average total variance ([1.072 – 1.043] × 100%). This is the contribution to total mean squared error due to the bias. In most cases, this contribution does not have such a noticeable impact.

Focusing on the means and medians of the 20 percent deletion rate, the results in Table 4.14 suggest that the gains from using REWT over LD in terms of bias reduction were roughly neutralized when looking at total mean squared error, because of the added variance of REWT, which uses more variable weights than LD.

The two hot-deck methods consistently produced lower total mean squared errors than LD and REWT. For some models, use of WSHD or CTBHD resulted in similar total mean squared errors; for others, CTBHD was clearly superior.

Often, the two maximum-likelihood methods were the best methods in terms of the resulting total mean squared errors. For one model (N19/ANLYR), however, neither worked as expected. MPLUS did not run at all, and LG produced relatively large total mean squared errors.

As expected, LD and REWT lose more information (have higher total variances and mean squared errors) than the other methods in most cases. Regardless of the model, the estimated total variance (Table 4.13) averages roughly 25 percent greater when using REWT than the estimated variance computed from the completed sample with a 20 percent deletion rate. That is exactly what is expected (1/.80 = 1.25). The average increase in the estimated total mean squared error (Table 4.14) is always higher, but it is less than 40 percent in every model. In contrast, the average increase in estimated total mean squared error when using the WSHD method peaks at 18 percent and is usually less than 10 percent.

Table 4.15 shows that the average bias in the variance estimator from using the WSHD method to estimate the N4/MHTRT model is 5.8 percent (–.058), which increases to 8.1 percent (–0.081) when viewed as an estimate of mean squared error in Table 4.16. Recall, all relative biases of estimated standard errors and root mean squared errors are roughly half the corresponding estimated variance and mean squared error.

Moreover, as expected, the variance estimator for REWT has systematically less bias than the variance estimators for the WSHD and CTBHD methods. REWT attempts to account for nonresponse, whereas the WSHD and CTBHD methods treat imputed values as if they were real and ignore the increase in variance due to imputation. The WSHD and CTBHD methods even underestimate the variance of the clone-dataset coefficients. This can be seen in Table 4.17, which displays the weighted distribution of the ratio of the variance estimate to the total variance for the coefficients computed from the completed sample for each method and deletion rate. The median and mean of these ratios are consistently (and erroneously) less than 1 for the WSHD and CTBHD methods, whereas they are greater than 1 for the other MIVRA methods.

Table 4.15 summarizes by model and MIVRA method the empirical relative biases of each method's variance *estimates* ($(e - v)/v$, with e from equation 4.9 and v from equation 4.7).

Because these variance estimates are the only measures available for mean squared error, Table 4.16 summarizes the estimated relative bias of variance estimates as measures of mean squared error ($(e - m)/m$, with m from equation 4.8). The bias in a method's estimate of standard error (and root mean squared error) is roughly half the bias in its estimate of variance (and mean squared error). The summaries in these two tables were computed using the square root of empirical total variance (Table 4.15) and mean squared error (Table 4.16) of the coefficient estimate as the weights.

Table 4.15 Empirical Relative Biases of Variance Estimates by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of Variance Estimates					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	-0.059	-0.040	-0.183	-0.421	-0.160	-0.097
	1st Quartile	0.008	-0.002	-0.016	-0.032	-0.010	-0.008
	Median	0.019	0.003	-0.010	-0.010	-0.008	0.001
	Mean	0.015	0.000	-0.024	-0.065	-0.022	-0.002
	3rd Quartile	0.033	0.007	-0.005	-0.005	-0.005	0.006
	Maximum	0.050	0.020	0.000	0.004	-0.001	0.113
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	-0.047	-0.072	-0.214	-0.025	-0.185	N/A
	1st Quartile	-0.013	-0.023	-0.053	-0.009	-0.027	N/A
	Median	0.001	-0.018	-0.044	-0.006	-0.020	N/A
	Mean	0.001	-0.018	-0.058	-0.007	-0.034	N/A
	3rd Quartile	0.012	-0.004	-0.037	-0.004	-0.017	N/A
	Maximum	0.048	0.001	-0.015	0.003	-0.001	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	-0.026	-0.033	-0.189	-0.036	-0.190	N/A
	1st Quartile	-0.009	-0.016	-0.138	-0.014	-0.132	N/A
	Median	0.000	-0.010	-0.017	-0.004	-0.017	N/A
	Mean	-0.001	-0.010	-0.050	-0.009	-0.050	N/A
	3rd Quartile	0.008	-0.002	-0.007	-0.001	-0.007	N/A
	Maximum	0.029	0.014	0.001	0.005	0.002	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	-0.193	-0.179	-0.215	-0.098	-0.251	N/A
	1st Quartile	-0.044	-0.051	-0.141	-0.015	-0.155	N/A
	Median	-0.019	-0.017	-0.018	-0.009	-0.021	N/A
	Mean	-0.033	-0.038	-0.056	-0.015	-0.062	N/A
	3rd Quartile	-0.001	-0.007	-0.011	-0.003	-0.009	N/A
	Maximum	0.038	0.018	0.009	0.002	0.011	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	-0.108	-0.044	-0.225	N/A	-0.230	-0.967
	1st Quartile	-0.060	-0.019	-0.152	N/A	-0.153	-0.967
	Median	-0.021	-0.010	-0.113	N/A	-0.100	-0.938
	Mean	-0.035	-0.012	-0.108	N/A	-0.103	-0.649
	3rd Quartile	-0.005	-0.003	-0.023	N/A	-0.016	-0.148
	Maximum	0.006	0.013	-0.011	N/A	-0.005	0.779

Table 4.15 Empirical Relative Biases of Variance Estimates by MIVRA Method for 20 Percent Deletion Rate (continued)

Study/Model	Summary Statistics	Empirical Relative Bias of Variance Estimates					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	-0.072	-0.079	-0.204	-0.072	-0.206	N/A
	1st Quartile	-0.044	-0.024	-0.129	-0.035	-0.167	N/A
	Median	-0.021	-0.018	-0.082	-0.014	-0.067	N/A
	Mean	-0.026	-0.021	-0.088	-0.022	-0.095	N/A
	3rd Quartile	-0.005	-0.008	-0.022	-0.004	-0.013	N/A
	Maximum	0.020	0.011	-0.008	0.003	-0.005	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across coefficients of $(e - v)/v$ from equations 4.7 and 4.9, with the coefficients' $\sqrt{v}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table 4.16 Empirical Relative Biases of MSE Estimates by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of MSE Estimates					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	-0.357	-0.437	-0.286	-0.904	-0.267	-0.196
	1st Quartile	-0.119	-0.065	-0.079	-0.858	-0.048	-0.058
	Median	-0.054	-0.040	-0.016	-0.164	-0.009	-0.007
	Mean	-0.078	-0.081	-0.049	-0.306	-0.045	-0.029
	3rd Quartile	-0.010	-0.009	-0.006	-0.011	-0.006	0.004
	Maximum	0.031	0.019	-0.003	0.004	-0.001	0.098
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	-0.366	-0.313	-0.347	-0.077	-0.208	N/A
	1st Quartile	-0.078	-0.049	-0.094	-0.010	-0.034	N/A
	Median	-0.044	-0.029	-0.047	-0.007	-0.021	N/A
	Mean	-0.065	-0.056	-0.081	-0.009	-0.038	N/A
	3rd Quartile	-0.024	-0.023	-0.040	-0.004	-0.018	N/A
	Maximum	0.042	0.000	-0.016	0.003	-0.002	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	-0.043	-0.037	-0.263	-0.050	-0.251	N/A
	1st Quartile	-0.014	-0.017	-0.140	-0.014	-0.135	N/A
	Median	-0.005	-0.011	-0.017	-0.004	-0.018	N/A
	Mean	-0.006	-0.012	-0.058	-0.011	-0.058	N/A
	3rd Quartile	0.005	-0.004	-0.007	-0.001	-0.007	N/A
	Maximum	0.025	0.014	0.001	0.005	0.002	N/A

Table 4.16 Empirical Relative Biases of MSE Estimates by MIVRA Method for 20 Percent Deletion Rate (continued)

Study/Model	Summary Statistics	Empirical Relative Bias of MSE Estimates					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	-0.202	-0.188	-0.215	-0.101	-0.251	N/A
	1st Quartile	-0.060	-0.051	-0.149	-0.015	-0.167	N/A
	Median	-0.035	-0.018	-0.018	-0.010	-0.022	N/A
	Mean	-0.045	-0.040	-0.060	-0.016	-0.065	N/A
	3rd Quartile	-0.004	-0.009	-0.013	-0.004	-0.013	N/A
	Maximum	0.037	0.017	0.009	0.002	0.011	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	-0.564	-0.385	-0.466	N/A	-0.458	-0.972
	1st Quartile	-0.213	-0.071	-0.226	N/A	-0.218	-0.968
	Median	-0.063	-0.017	-0.170	N/A	-0.130	-0.952
	Mean	-0.116	-0.057	-0.181	N/A	-0.157	-0.752
	3rd Quartile	-0.035	-0.007	-0.085	N/A	-0.070	-0.791
	Maximum	0.002	0.003	-0.016	N/A	-0.005	0.236
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	-0.137	-0.091	-0.264	-0.103	-0.249	N/A
	1st Quartile	-0.068	-0.027	-0.152	-0.035	-0.175	N/A
	Median	-0.040	-0.019	-0.100	-0.014	-0.085	N/A
	Mean	-0.046	-0.025	-0.104	-0.026	-0.105	N/A
	3rd Quartile	-0.017	-0.010	-0.023	-0.004	-0.014	N/A
	Maximum	0.010	0.007	-0.009	0.002	-0.007	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: Weighted distributions are across coefficients of $\left(e-m\right)/m$ from equations 4.8 and 4.9, with the coefficients' $\sqrt{m}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table 4.17 Ratios of Estimated Variance to Full-Sample Variance by MIVRA Method for 20 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Estimated Variance to Full-Sample Variance					
		LD	REWT	WSHD	MPLUS	CTBHD	LG
N4/SPDMON	Sample Size	32	32	32	32	32	32
	Minimum	1.112	1.124	0.951	0.686	0.981	0.960
	1st Quartile	1.198	1.256	0.992	0.983	0.996	0.999
	Median	1.238	1.278	0.996	0.997	0.998	1.005
	Mean	1.234	1.267	0.995	0.967	0.997	1.008
	3rd Quartile	1.264	1.296	0.998	1.003	0.999	1.008
	Maximum	1.373	1.365	1.020	1.011	1.014	1.138

Table 4.17 **Ratios of Estimated Variance to Full-Sample Variance by MIVRA Method for 20 Percent Deletion Rate (continued)**

Study/Model	Summary Statistics	Ratio of Estimated Variance to Full-Sample Variance					
		LD	**REWT**	**WSHD**	**MPLUS**	**CTBHD**	**LG**
N4/MHTRT	Sample Size	34	34	34	34	34	N/A
	Minimum	1.135	1.148	0.963	1.000	0.989	N/A
	1st Quartile	1.192	1.221	0.974	1.005	0.997	N/A
	Median	1.219	1.231	0.981	1.007	1.000	N/A
	Mean	1.222	1.236	0.982	1.015	1.001	N/A
	3rd Quartile	1.249	1.266	0.984	1.013	1.003	N/A
	Maximum	1.352	1.331	1.008	1.114	1.017	N/A
N14/YOTMTHLP	Sample Size	35	35	35	35	35	N/A
	Minimum	1.145	1.140	0.916	0.998	0.917	N/A
	1st Quartile	1.170	1.165	0.976	1.001	0.973	N/A
	Median	1.183	1.186	0.992	1.003	0.994	N/A
	Mean	1.191	1.188	0.985	1.015	0.985	N/A
	3rd Quartile	1.200	1.203	0.998	1.021	0.997	N/A
	Maximum	1.281	1.250	1.006	1.061	1.006	N/A
N14/YORXHLP	Sample Size	34	34	34	34	34	N/A
	Minimum	0.976	1.086	0.867	0.978	0.840	N/A
	1st Quartile	1.142	1.151	0.961	0.997	0.956	N/A
	Median	1.165	1.172	0.989	1.001	0.989	N/A
	Mean	1.161	1.168	0.980	1.012	0.977	N/A
	3rd Quartile	1.187	1.182	0.998	1.007	0.998	N/A
	Maximum	1.282	1.276	1.016	1.115	1.015	N/A
N19/ANLYR	Sample Size	26	26	26	N/A	26	26
	Minimum	1.117	1.186	0.907	N/A	0.904	0.954
	1st Quartile	1.136	1.212	0.941	N/A	0.944	1.089
	Median	1.175	1.239	0.954	N/A	0.960	1.225
	Mean	1.187	1.241	0.958	N/A	0.964	6.404
	3rd Quartile	1.226	1.265	0.985	N/A	0.992	1.277
	Maximum	1.352	1.386	0.994	N/A	0.999	131.760
N19/ABODANL	Sample Size	26	26	26	26	26	N/A
	Minimum	1.102	1.164	0.919	0.993	0.913	N/A
	1st Quartile	1.123	1.172	0.954	1.004	0.928	N/A
	Median	1.203	1.216	0.977	1.016	0.985	N/A
	Mean	1.187	1.217	0.971	1.018	0.967	N/A
	3rd Quartile	1.222	1.251	0.989	1.022	0.994	N/A
	Maximum	1.283	1.301	1.000	1.053	1.005	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = (pseudo-)maximum likelihood method using off-the-shelf Latent GOLD software; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

4.3.3 Statistical Significance of Regression Coefficients

When fitting a regression model, a key consideration for the analyst is whether an estimated coefficient is statistically significant. Missing data can adversely affect an analyst's ability to assess whether a regression coefficient is statistically significant due either to bias, a loss of statistical efficiency (increase in variance), or both.

This section reports the results of an investigation into how often a determination of statistical significance at the (two-sided) 5 percent level changes when using a particular method for handling missing item data rather than the (unknown in practice) completed sample. The missing data methods are compared in Table 4.18 under the 20 percent deletion rate. Results for the other deletion rates (Tables F.15 and F.16 in Appendix F) are analogous but, as expected, not as strong. This is taken into consideration in Chapter 5 when providing guidance on how to take missing data into account for regression analyses.

The rows in Table 4.18 denote the significance level based on the clone dataset, and the columns denote the significance level using the given missing data methods. The individual tables summarize the performance of a missing data method across all covariates and all simulation runs.

A method that performs well would agree with the clone-dataset significance results practically all the time. What little disagreement there is should occur mostly in the direction of *less* significance, because missingness means less data and therefore less power. False negatives (i.e., failure to detect statistical significances that exist in the completed sample) err on the conservative side. False positives (i.e., detecting statistical significances that do not exist in the completed sample) are more troubling because they allow the analyst to draw incorrect conclusions about associations. By contrast, it is widely accepted that the absence of evidence—here failing to detect statistical significance—should not be treated as evidence of absence of an association.

The expectation was that the two maximum-likelihood methods and the two hot-deck methods should match the clone dataset almost all the time, but that the two hot-deck methods should exhibit more of the troubling type of error (i.e., the false positives), because the hot-deck methods do not account for the increased variance due to the item missingness. The LD and REWT methods should show the less troubling form of error more often than the other methods because the reduction in sample size should increase the estimated variance of the coefficient estimates.

The actual results in Table 4.18 met these expectations in some ways but not others. The MPLUS, CTBHD, and WSHD methods matched the clone-dataset significance results more often than did the LD and REWT methods. The LD and REWT methods tended to miss more often on the conservative side than the liberal side, and the other methods often missed more often on the liberal side than the conservative side. By far the most important factor influencing the results was the location of the clone-dataset *p*-values (i.e., whether they were close to the cut-off of 0.05 used in this study). Missingness in the covariates did not seem to be an important factor.

To give the reader a glimpse of how these results look at the covariate level, a more detailed discussion of one of the analytic models at the 20 percent deletion rate follows. Table 4.19 summarizes the results for the N19/ANLYR model. This model has the largest sample size and few covariates with clone-dataset *p*-values near 0.05. Out of the 26 covariates, 19 had clone-dataset *p*-values of less than 0.01. For practically all iterations of the simulation, each of the methods detected statistical significance for these. Three of the remaining seven had clone-dataset *p*-values greater than 0.25, and significance was not detected using any of the methods most of the time. For the four covariates with *p*-values between 0.01 and 0.25, some differences occurred between the methods. Some examples follow:

- The coefficient for the covariate "Family income $20,000-$49,999" had a clone-dataset *p*-value of 0.0108. Ideally, the MIVRA methods examined would detect significance here most or all the time, regardless of the method and number of iterations. WSHD, CTBHD, and LG detected significance for every iteration, but LD failed to detect significance for 507 of the 1,600 iterations (31.7 percent), and REWT failed to detect significance for 551 of the 1,600 iterations (34.4 percent). This is an example of the effect of higher variances associated with the LD and REWT methods because of the reduction in sample size.
- The coefficient for the covariate "Rural" had a clone-dataset *p*-value of 0.1002. WSHD, CTBHD, and LG did not detect significance for any of the iterations, but LD detected significance for 104 of the 1,600 iterations (6.5 percent), and REWT detected significance for 190 of the 1,600 iterations (11.9 percent). This is an example of the relative instability of the LD and REWT methods, something that was observed across all covariates and deletion rates. Although these methods tend to miss on the conservative side more than the liberal side, they also tend to miss more often overall. The induced missingness might affect the *p*-values in either direction; the *p*-values move frequently, depending on which 20 percent of the records are discarded. For the other methods, no records are being discarded. (This refers to the distinction between the deletion rate and the overall covariate missingness rate.)
- The coefficient for the covariate "6-24 religious services in the past year" had a clone-dataset *p*-value of 0.0454. It would be expected that the LD and REWT methods would fail to detect significance somewhat frequently because the *p*-value is barely below $\alpha = 0.05$. LD fails for 654 of the 1,600 iterations (40.9 percent), which is not much more frequently than WSHD and CTBHD. However, REWT fails for 1,022 of the 1,600 iterations (63.9 percent). This is likely because REWT has more variance than LD due to the weight adjustments.

Analogous tables for the other five analytic models tell a similar story (20 percent only) and are displayed in Tables F.17 to F.21 in Appendix F. Differences across the methods are often more noticeable because there are more covariates with clone-dataset *p*-values near 0.05. In numerous cases, either covariates with clone-dataset *p*-values just below 0.05 are undetected by LD and REWT or covariates with clone-dataset *p*-values just above 0.05 are detected by all the methods.

Table 4.18 Statistical Significance of Covariates by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

	LD		REWT		WSHD		MPLUS		CTBHD		LG	
Study/Model	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**
N4/SPDMON												
α = 0.05 (23 coefficients)	69.9	2.0	70.7	1.2	71.8	0.0	71.8	0.1	71.9	0.0	71.9	0.0
N.S. (9 coefficients)	1.5	26.6	2.3	25.8	0.2	27.9	0.6	27.5	0.1	28.0	0.0	28.1
N4/MHTRT												
α = 0.05 (12 coefficients)	28.8	6.5	28.6	6.7	34.3	1.0	34.9	0.4	34.4	0.9	N/A	N/A
N.S. (22 coefficients)	4.4	60.3	4.1	60.6	3.6	61.1	1.7	63.0	2.3	62.4	N/A	N/A
N14/YOTMTHLP												
α = 0.05 (13 coefficients)	29.5	7.6	29.6	7.5	35.3	1.8	35.7	1.5	35.1	2.1	N/A	N/A
N.S. (22 coefficients)	2.2	60.7	2.3	60.6	1.5	61.4	0.7	62.2	1.4	61.5	N/A	N/A
N14/YORXHLP												
α = 0.05 (9 coefficients)	20.4	6.1	20.3	6.1	22.7	3.8	23.4	3.1	23.2	3.3	N/A	N/A
N.S. (25 coefficients)	3.4	70.1	3.8	69.7	2.1	71.5	1.6	71.9	2.2	71.3	N/A	N/A
N19/ANLYR												
α = 0.05 (21 coefficients)	73.7	7.0	72.4	8.4	79.6	1.2	N/A	N/A	79.4	1.4	77.2	3.5
N.S. (5 coefficients)	1.7	17.6	1.8	17.5	1.5	17.7	N/A	N/A	0.7	18.6	0.0	19.2
N19/ABODANL												
α = 0.05 (11 coefficients)	37.4	4.9	35.8	6.5	37.4	4.9	40.1	2.2	37.6	4.8	N/A	N/A
N.S. (15 coefficients)	2.4	55.3	1.9	55.8	1.6	56.1	1.6	56.1	1.7	56.0	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = Latent GOLD; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software applied to all variables in the model being fit; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite. The LG method was implemented only for two models at the 20 percent deletion rate (Section 4.2.4).

Note: See Table 4.1 for study and model details.

Table 4.19 Statistical Significance of Individual Covariates in N19/ANLYR Model by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		CTBHD		LG	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Had Past Year Major Depressive Episode											
Yes	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Support											
At Least 3	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
4-5	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
6-7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age Group											
12-13	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
14-15	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
16-17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Gender											
Male	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity											
White	0.0022	1,599	1	1,590	10	1,600	0	400	0	49	0
Black/African American	0.0708	585	1,015	534	1,066	619	981	69	331	0	49
Other	0.9080	0	1,600	0	1,600	0	1,600	0	400	0	49
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Income											
Less than $20,000	0.0044	584	1,016	588	1,012	1,600	0	400	0	49	0
$20,000-$49,999	0.0108	507	1,093	551	1,049	1,600	0	400	0	49	0
$50,000-$74,999	0.2786	7	1,593	4	1,596	0	1,600	0	400		
$75,000 or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rural/Urban Segment											
Rural	0.1002	104	1,496	190	1,410	0	1,600	0	400	0	49
Urban	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Alcohol Use Disorder											
Yes	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table 4.19 **Statistical Significance of Individual Covariates in N19/ANLYR Model by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)**

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		CTBHD		LG	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Past Year Illicit Drug Use Disorder Excluding Pain Relievers											
Yes	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Number of Delinquent Behaviors											
None	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
One	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
Two or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Grades for Last Semester											
An "A+," "A," or "A-Minus" Average	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
A "B+," "B," or "B-Minus" Average	0.0000	1,600	0	1,600	0	1,600	0	400	0	49	0
A "C+," "C," or "C-Minus" Average	0.0001	1,600	0	1,600	0	1,600	0	400	0	49	0
A "D" or Less than a "D" Average	0.0048	1,545	55	1,495	105	1,577	23	396	4	49	0
School Does Not Give These Grades	0.0000	1,600	0	1,599	1	1,600	0	400	0	49	0
Not Enrolled in School	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
How Many Religious Services Attended in Past Year											
0 Times	0.0020	1,586	14	1,534	66	1,600	0	400	0	49	0
1-5 Times	0.0062	1,508	92	1,376	224	1,599	1	400	0	49	0
6-24 Times	0.0454	946	654	578	1,022	1,118	482	260	140	4	45
25-52 Times	0.3986	3	1,597	2	1,598	0	1,600	0	400	0	49
More than 52 Times	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = Latent GOLD; MIVRA = missing item values in regression analyses; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. These rows correspond to reference levels of covariates.

Note: See Table 4.1 for study and model details.

4.4 Discussion: Impact of the Missingness of the Independent and Dependent Variables

One of the surprising results of the simulation study was the success of the WSHD method. This simple and popular method did not underestimate variance to the extent expected (Table 4.15), and despite the use of only a few auxiliary variables, WSHD corrected for bias reasonably well (Tables 4.11 and 4.12). In fact, it was the surprising performance of WSHD that led to the addition of the more sophisticated imputation method, CTBHD.

One possible explanation for the performance of WSHD may be due to the true amount of missing data. There are various ways to measure the amount of missing data. Tables 2.4 through 2.9 display the actual missingness rates of the independent variables in the six models under investigation, and Table 2.3 displays the deletion rates for these six models, with and without zero-fill imputation. One thing these tables do not show is the fraction of independent variables that are missing, termed "X-matrix missingness." This represents the total number of cells with missing data in the matrix created by all the cases and variables used in the analysis. That can be found in Table 4.20, which also displays the actual fraction that would be missing with overall 5, 12.5, and 20 percent deletion rates. The larger the difference between the deletion rate and this fraction, the more information is lost from using the two listwise deletion methods, LD and REWT. Imputation methods like WSHD and CTBHD might be expected to perform well when the X-matrix missingness is low, because there are not many missing values to replace (reducing the potential for bias), and all the nonmissing values are preserved. Because values that were imputed using the complex NSDUH imputation method were treated as nonmissing, and because many of the covariates in the model were of this type, the X-matrix missingness is fairly low for all six models.

Table 4.20 Missingness of Independent Variables

Study/Model	Sample Size	X-Matrix Missingness by Deletion Rate				Is Deletion a Function of the Dependent Variable?	
		Actual	5%	12.5%	20%	Test Statistic	*P*-Value
N4/SPDMON	93,121	0.27	0.31	0.77	1.24	0.36	0.72
N4/MHTRT	7,609	0.31	0.33	0.82	1.32	1.13	0.26
N14/YOTMTHLP	3,308	1.80	0.57	1.42	2.28	1.20	0.31
N14/YORXHLP	1,545	2.05	0.60	1.50	2.40	0.37	0.83
N19/ANLYR	112,591	2.61	1.00	2.50	4.00	-7.04	0.00
N19/ABODANL	7,084	2.08	0.82	2.06	3.30	-2.23	0.03

Note: The last two columns are the results of regressing the deletion indicator on the dependent variables and all independent variables that are never missing. The two N14 models use Wald F tests rather than *t* tests because the dependent variable is multivariate.
Note: See Table 4.1 for study and model details.

The table also shows the results of testing whether the dependent variable is significant when a 0/1 deletion indicator (i.e., the indicator is 1 when any independent variable has a missing value leading to the record's removal in listwise deletion) is regressed on the dependent variables and all the independent variables that are never missing. The absolute biases from using LD as opposed to REWT should be greater when listwise deletion is a function of the dependent variable even when holding constant all the independent variables that are never missing. This was *not* borne out in the analysis (shown in Table 4.11), in large part, because the estimated relationship between a record's deletion and the never-missing variables (including the model's dependent variable) was simulated as the true relationship regardless of the size of the test statistic. Thus, the LD method was *nearly always* the more biased of the two. Surprisingly, it rarely had more mean squared error (Table 4.14), even when the bias was "statistically significant" (*p*-values less than 0.05). However, the bias of the mean squared error measure was much larger when the LD method was used in those cases.

This page intentionally left blank

5. Recommendations and Conclusions

5.1 Introduction

This chapter provides recommendations to analysts about whether and when to use listwise deletion (LD), LD with reweighting (REWT), two methods of hot-deck imputation, and two software packages using maximum-likelihood (ML) methods that account for the nature of a complex survey design. These recommendations are based on the literature review on several methods for handling missing item values in regression analyses (MIVRA) described in Chapter 3 and the additional insights that were gleaned from the simulation experiment on six National Survey on Drug Use and Health (NSDUH) datasets described in Chapter 4.

No multiple imputation (MI) methods are included in these recommendations because, as discussed previously, standard MI methods do not fully account for the complex survey design except under very strong assumptions (Kim, Brick, Fuller, & Kalton, 2006). Moreover, each multiply imputed dataset differs and does not allow for exact replicability of results. Therefore, MI may be a technique that is best used by the producers of the dataset rather than by those analyzing it for consistency in the resulting analyses.

The method of treating missing item values for an independent variable as an additional level or category of that variable is also not recommended. The literature review in Chapter 3 showed that this is not an appropriate MIVRA option (Jones, 1996) unless the category of missing is of analytic interest (because, for example, the analyst is interested in estimating the fraction of the population who cannot answer a particular question).

One fact to keep in mind is that in all the simulations, the estimates of regression coefficients exhibited statistically significant biases using every one of the MIVRA methods evaluated. These biases, however, were usually not very meaningful; that is, the fraction of a regression coefficient's (empirical) mean squared error that was attributable to bias was often quite small. The reason for the relatively small, but statistically significant, biases is likely that the assumptions underlying the analytic model being fit do not hold (e.g., for a binary logistic model, there is no $\boldsymbol{\gamma}$ such that $\mathrm{E}\left(y\right) = 1/[1 + \exp(\mathbf{x}^T\boldsymbol{\gamma})]$ no matter what the value of $\mathbf{x}$ is). This is likely a situation that is often faced in a regression analysis where the hypothesized model comes close enough to holding to be scientifically useful[23] but does not, strictly speaking, hold throughout the population of interest. There are no hard and fast rules for determining when that occurs, and the decision about the usefulness of a model is often left up to the analyst who presumably knows the data and subject matter.

Another factor is that when the value of the dependent variable was missing from a record or it could not be determined whether the sampled record was in the subpopulation of interest, it was dropped from the analysis.[24] Only alternative methods of treating missing

[23] As Box and Draper (1987, p. 74) put it, "All models are wrong; the practical question is how wrong do they have to be to not be useful."

[24] Both occurrences were rare in NSDUH data. Missing dependent variables, when they occurred, were almost always imputed using a sophisticated method that is assumed to not lead to biases.

independent variable values were evaluated. This assumes that the dependent variable's missingness is a function of the independent variables in the regression model, a commonly made, but not always correct, assumption. When it fails, parameter estimates will likely be biased. Section 5.4 includes a discussion about what can be done when the dependent variable is missing.

Section 5.2 summarizes the methods that were evaluated as well as their strengths and weaknesses. Section 5.3 provides analysts with decision trees that are intended to help analysts identify the optimal MIVRA method for their specific analytic conditions. Finally, Section 5.4 provides concluding remarks and thoughts about future directions and research into MIVRA methods for complex survey data.

5.2 Evaluation of MIVRA Methods

The LD method drops all records from a regression analysis that have any missing covariates. Its use can lead to biases that meaningfully impact the results when the probability that a record has been deleted from the analysis is a function of the dependent variable after accounting for the values of the independent variables.[25] This is something the user can evaluate by, for example, running a logistic regression with the dependent variable as a 0/1 indicator of whether or not a record is deleted while the independent variables include the original dependent variable of the regression model being analyzed and other variables that are never missing (Section 4.4). This again assumes that the dependent variable itself is never missing.

A second problem with LD is that it removes what may be useful information, because even if a value is missing from only one independent variable, then the entire record is deleted. Even when there is not a meaningful bias from the deletion of whole records, the percentage increase in the standard errors of regression coefficients due to LD is roughly half the percentage of records being deleted when the percentage of deleted records is less than 50 percent.

The REWT method deletes the same records as LD, but it uses estimates of the probabilities that records with particular characteristics remain in the listwise-deleted dataset to adjust their complex sampling weights (the weight of a record in the listwise-deleted dataset is multiplied by the inverse of the estimated probability that it has not been deleted from the set). This method usually reduces the bias of LD, if such a bias exists, but shares the problem that it can remove much useful information. Again, even when there is not a meaningful bias, the percentage increase in the standard errors of regression coefficients due to REWT is roughly half the percentage of records being deleted. In fact, it is often higher than that of LD, so much so, that even with reduced bias, the mean squared error for using REWT is often larger than that from using LD.

Although LD and REWT can produce relatively large standard errors, compared with the alternatives to be discussed later, the standard errors computed after using the two LD methods have relatively small biases. Thus, these methods often produce useful inferences, if not as accurate (i.e., estimated coefficients have larger mean squared errors) as those of the alternatives.

[25] See footnote in Section 3.2.1 for a discussion of why the terms "missing at random" and "missing not at random" are not used here.

In addition to their simplicity, an advantage of LD and REWT over most other MIVRA methods is that their properties are not affected when the probability of a record being deleted is a function of the covariate values that are missing, so long as the regression model being analyzed is correct in the population of interest (Little, 1992; Kott, 2015). This advantage was not evaluated in the simulations; however, these issues were considered when providing recommendations.

In donor imputation, the item value of a respondent with characteristics similar to the item nonrespondent is used in place of the latter's missing value. Donor imputation, unlike LD (with or without reweighting), uses all the respondent's available information. In principle, the method can be subject to bias if the donor values are not truly representative of the missing values.

Two versions of donor imputation were implemented on the six NSDUH datasets. Both create imputation cells and then use a weighted sequential hot deck (WSHD) to choose the donors within cells. The first version created cells using gender and race/ethnicity and then sorted by age for WSHD. The second version used regression trees with all the variables in the regression model being analyzed as potential predictors to create the cells and then sorted the members of each cell by gender, race/ethnicity, and age for WSHD (more details are in Chapter 5). Not surprisingly, the method using the more involved approach to creating imputation cells tended to produce regression coefficient estimates with less biases. These biases were comparable with those from REWT. With the NSDUH datasets analyzed, however, the added biases from using the simpler donor imputation method usually were not very large. That may be a specific characteristic of NSDUH rather than a hard and fast rule about using imputation cells and sorting based only on gender, race/ethnicity, and age. With other datasets, using gender, race/ethnicity, and age alone to form cells and sort within them may not be as successful.

The larger problem with the donor imputation methods in general is that, although often reliable in terms of bias and actual mean squared error, they almost always underestimate standard errors because they treat imputed values as if they were true values. Users of donor imputation should be aware of this and interpret their results with caution. The larger the amount of missingness, and therefore the greater the percentage of imputed values used, the more the need for caution.

Two software packages that use sophisticated (pseudo-)maximum-likelihood methods to address missing covariate item values in a regression analysis performed on complex survey data were examined: Mplus® and Latent GOLD®. When they work, these programs usually give the best results in terms of bias, mean squared error, and standard error estimation, which are strong reasons for a user to consider employing them. However, in addition to requiring the user to purchase software he or she may not already possess, Mplus could not handle the two larger datasets,[26] and Latent GOLD ignored the variance contribution from (variance) strata containing records in the subpopulation of interest (e.g., pregnant women) from only one (variance) primary sampling unit (PSU). Latent GOLD, unlike Mplus and SUDAAN® (RTI International, 2013), does not recognize that there were sampled records not in the subpopulation of interest from

[26] For one of the datasets with large sample size, Mplus produced poor results, and for the other, Mplus reported that it lacked the memory to fit the model.

other PSUs of such strata. Therefore, the variance contribution of these strata is not accounted for by Latent GOLD.[27]

As noted before, a problem with all these methods is that they assume the regression model fit holds in the population of interest. ML methods may be the most sensitive to the failure of this assumption. REWT, by contrast, may not be as sensitive. This results when reweighting accounts, at least approximately, for the true probabilities of a record being deleted.

5.3 Recommended MIVRA Methods for Specific Analyses

In this section, MIVRA methods are recommended for specific analyses using decision trees. Before choosing a MIVRA method, however, an analyst should consider the following questions.

- *What is the deletion rate in the complex sample being analyzed?* The LD method is not recommended for a deletion rate (fraction of records deleted if a single variable has a missing value) greater than 10 percent because of its limited ability to correct for nonresponse bias.
- *What percentage of the independent variable values are missing?* This is called "X-matrix missingness" here and in Section 5.6 (i.e., in the complete matrix of independent variable values, with the variables serving as columns and the records serving as rows, what fraction of values are missing).
- *Is the missingness in the independent variables (the X-matrix) a function of the dependent variable (y) of the analytic (regression) model being fit?* The test for this is mentioned in Section 4.4.
- *Do the schedule and budget allow for a sophisticated method?* Even for analysts comfortable with hot-deck imputation methods and ML methods, LD and REWT are the easiest methods to use.
- *How many of the independent variables have no missing values?* The REWT method works best when there are several independent variables with no missing values.
- *Is there interest in estimating the sizes of the regression coefficients of the model or only in determining whether they are statistically significantly different from zero?* The latter is called "Statsig only" in the decision trees. It is often of interest in preliminary analyses of the large datasets with many potential independent variables.
- *If there is interest only in determining whether the regression coefficients are statistically significantly different from zero, are there any coefficients with p-values slightly greater than* α *under LD?* This issue is discussed in Section 5.7. Donor imputation methods tend to underestimate variance, causing the *p*-values of the regression coefficients to be too small. These methods may find significance for

[27] Latent GOLD 5.0 (the version used in this study) ignores the variance contribution from single-unit strata. This is not a major issue for large NSDUH sample sizes, because most or all variance PSUs will likely be represented in the subsample. For example, of the six models in Table 4.1, N4/SPDMON and N19/ANLYR have representation in all 1,800 variance PSUs. N4/MHTRT and N19/ABODANL have 39 and 65 single-unit strata, respectively, and N14/YOTMTHLP and N14/YORXHLP have 258 and 440 single-unit strata, respectively. Note that Version 5.1 of the software allows the same options as Stata for variance estimation in the presence of single-unit strata.

independent variables when they should not (false positives). If no *p*-values under LD are slightly greater than α, then the risk of false positives is greatly diminished.[28,29]

- *Is the analyst comfortable with ML methods?* These methods have much appeal when the analyst is confident in the analytic model being fit, in which case they are recommended for practically all MIVRAs where they can be applied (recall that they sometimes fail).
- *How many strata in the dataset have only one responding sampling unit and how many have more than one? If there are some single-unit strata, is the analyst comfortable with aggregating strata? Are there too many single-unit strata to make this feasible?* Version 5.0 of Latent GOLD does not account for variance from strata with only one responding sampling unit. This causes the underestimation of variance and increases the risk of finding statistically significant covariates that should be treated as not statistically significant unless strata are collapsed together in some manner (which can overestimate variances).
- *How many observations does the model have (including the ones with one or more missing independent variables)? Is the sample size larger than 25,000?* In the simulation experiment, MPLUS had trouble with models with large sample size.

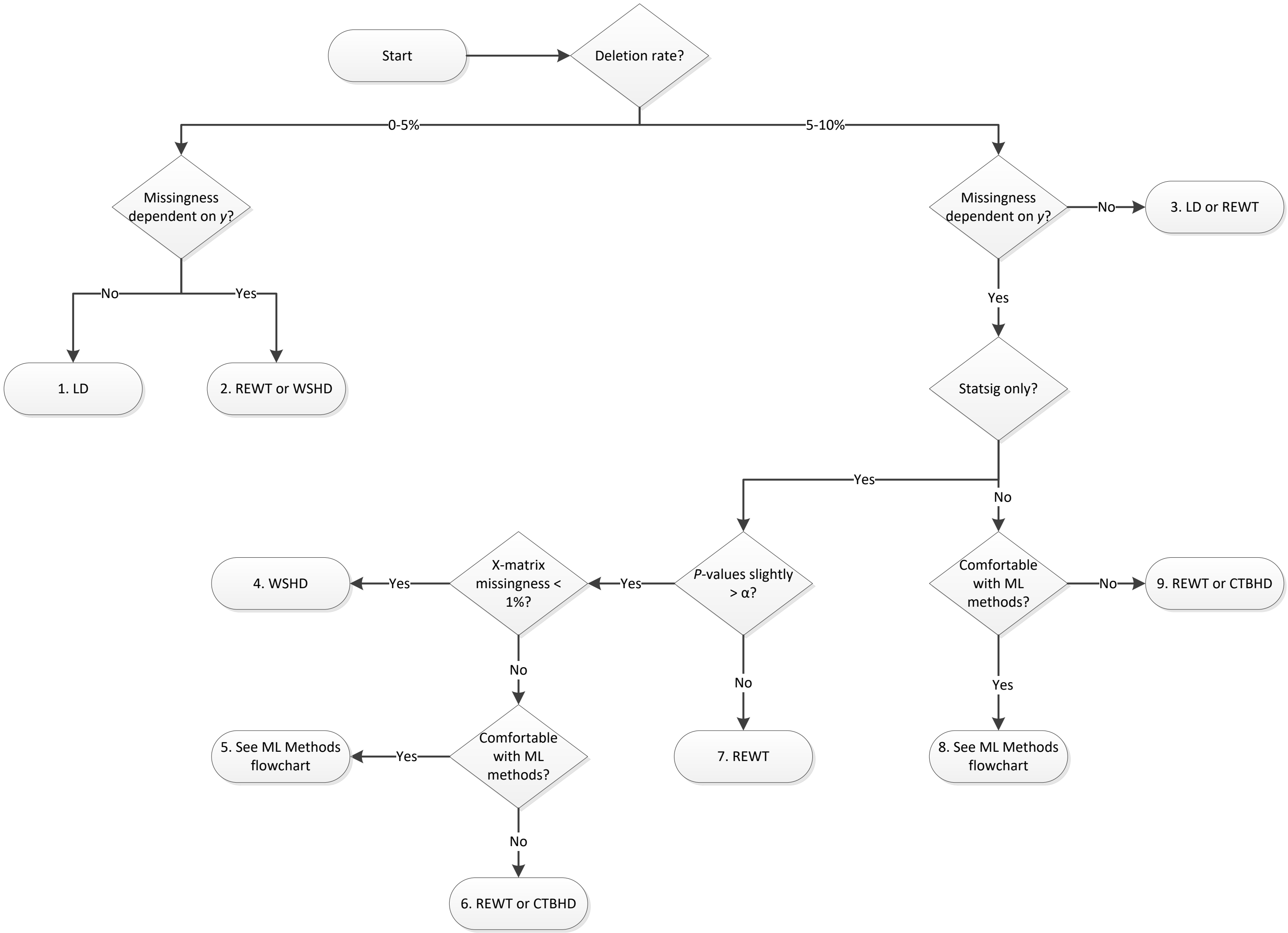
5.3.1 Deletion Rates Less than 10 Percent

A decision tree for deletion rates less than 10 percent is presented in Exhibit 5.1. The oval shapes denote the final leaves, or terminal nodes, in the tree. Each terminal node is numbered. A more detailed justification for the decisions follows the exhibit. When an ML method is selected, it can only be used when the procedure does not fail because the dataset is too large (MPLUS) or when there are at least two PSUs from each stratum in the dataset being analyzed (Latent GOLD).

[28] Note that in some scenarios, false negatives are considered at least as risky as false positives. For example, if the independent variable describes some public health intervention, it may be considered riskier to fail to intervene when one should intervene instead of intervening when it is unnecessary.

[29] The discussion in Section 5.7 is based on $\alpha = 0.05$, but the general conclusions in that section are expected to apply to other values of α as well. What constitutes "slightly greater" than α is for the analyst to decide. It depends on the sample size and the number of coefficients in a model. If the analyst has the resources, using more than one MIVRA method to see if the same significance result is reached is recommended.

Exhibit 5.1 Decision Tree for Deletion Rates Less than 10 Percent



CTBHD = cyclical tree-based hot deck; LD = listwise deletion; ML = maximum-likelihood; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

Note: The ML Methods flowchart is presented in Exhibit 5.4.

Terminal Node 1. With a low deletion rate and missingness that does not depend on *y*, there should be little risk for bias, and the variance of the regression coefficients should not increase by much using LD.

Terminal Node 2. If the missingness depends on *y*, even for a low deletion rate, a simple MIVRA method like REWT or WSHD is recommended to correct for any nonresponse bias.

Terminal Node 3. Even for a deletion rate in the 5 to 10 percent range, if the missingness is not dependent on *y*, nonresponse bias should be somewhat ignorable. LD should be fine, but REWT may better correct for any potential biases due to a misspecified analytic model.[30] The WSHD and cyclical tree-based hot-deck (CTBHD) methods are not recommended in this case, because even WSHD methods depend on the model being correctly specified, and LD and REWT are not discarding an inordinate amount of data.

When the analysis is such that only statistical significance and not the estimated coefficient is the focus, terminal nodes 4 to 7 apply to analyses with deletion rates in the 5 to 10 percent range in which missingness depends on *y*. For these analysts, LD is not recommended because the missingness depends on *y*, and this is easily addressed by REWT. The main concerns are that (1) REWT will fail to identify significant independent variables (false negatives), and (2) WSHD and CTBHD will find significance for independent variables when they should not (false positives). The false positives are more of a concern than the false negatives. False positives are likeliest to occur for covariates whose *p*-values under LD are slightly greater than α, because covariates whose *p*-values under LD are slightly greater than α might also have true (full-sample) *p*-values slightly greater than α. If the true *p*-value is slightly greater than α (say, 0.06 for $\alpha = 0.05$), an MIVRA method that tends to underestimate variance might have a high probability of identifying it as statistically significant.

Terminal Node 4. Even though the deletion rate is in the 5 to 10 percent range, not many missing data are in the X-matrix. WSHD is unlikely to underestimate the variance by much, so it is the recommended method. REWT would be discarding a fair number of records when that seems unnecessary. CTBHD and ML are also possibilities, but they would require more time and effort for most analysts and are unlikely to produce markedly different results than WSHD due to the low X-matrix missingness.

Terminal Node 5. Generally, ML methods are recommended whenever the deletion rate is greater than 5 percent, the missingness is dependent on *y*, and the X-matrix missingness is greater than 1 percent.

---

[30] Kott (2007) provides an interpretation of what it means to estimate the parameters of a model that does not fit the population. When the analytic model $y_k = \mathrm{f}(\mathbf{x}_k{}^T\boldsymbol{\beta}) + \varepsilon_k$ fits the population of interest, the conditional expectation is that ε_k is 0 no matter what the value of $\mathbf{x}_k$. This requirement of the standard model rarely ever holds. By contrast, the looser requirement for what Kott calls the "extended model" in which ε_k has an unconditional expectation in the population of 0 and is uncorrelated with the components of $\mathbf{x}_k$ almost always does. REWT has a greater potential for fitting the extended model when the standard model fails than LD.

Terminal Node 6. With more missing data in the X-matrix, WSHD might be underestimating variance more severely, causing more false positives. REWT is less likely to underestimate variance, making it the safer choice. CTBHD should be better than WSHD and at least as good as REWT at addressing bias, perhaps counterbalancing its higher risk (compared with REWT) of false positives. If ML methods are unavailable or fail, REWT or CTBHD are recommended.

Terminal Node 7. If there are few or no covariates with *p*-values slightly greater than α, REWT is unlikely to result in too many false negatives.

Terminal Node 8. See terminal node 5.

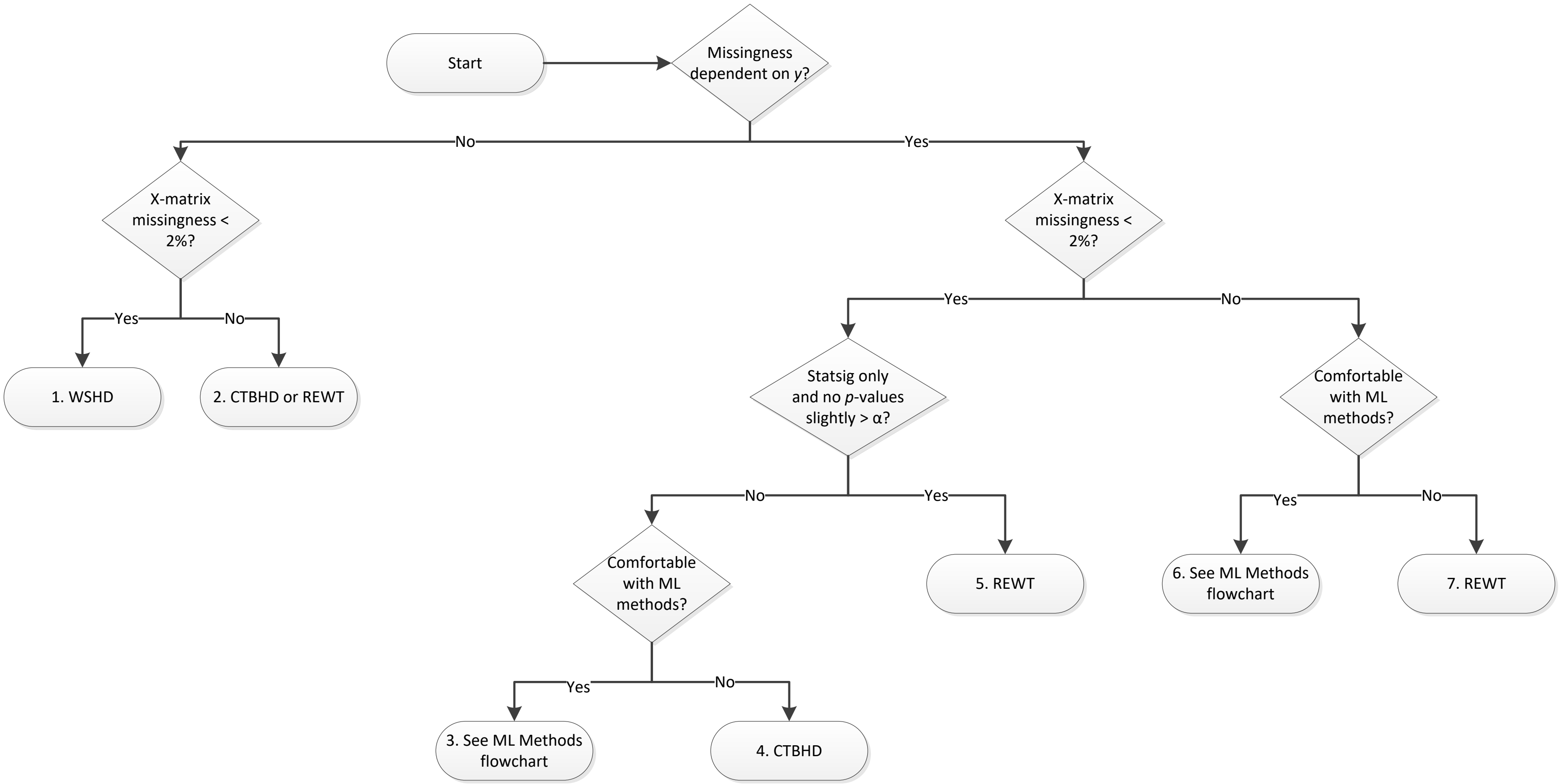
Terminal Node 9. REWT is a better choice for analysts concerned with false positives, whereas CTBHD is a fair choice for those more concerned about discarding too many records. With the deletion rate no greater than 10 percent, the underestimation of variance by CTBHD should not be too severe. Similar to terminal node 6, REWT or CTBHD are recommended if ML methods are unavailable or fail.

Terminal nodes 8 and 9 are for analysts who are facing a deletion rate in the 5 to 10 percent range in which missingness depends on *y* and who are interested in the sign and magnitude of the regression coefficients. For these analysts, LD is not recommended because the missingness depends on *y*, and WSHD is not recommended because there is more of a risk of bias in the regression coefficients. In the simulation experiment, REWT, CTBHD, and ML all performed better for bias than LD and WSHD.

5.3.2 Deletion Rates between 10 and 15 Percent

The decision tree for deletion rates between 10 and 15 percent is presented in Exhibit 5.2. A discussion of the decisions follows the exhibit. Again, when an ML method is selected, it can only be used when the procedure does not fail because the dataset is too large (MPLUS) or when there are at least two PSUs from each stratum in the dataset being analyzed (Latent GOLD). The LD method is not recommended for deletion rates in this range due to its limited ability to correct for nonresponse bias, as mentioned at the beginning of Section 4.4.

Exhibit 5.2 Decision Tree for Deletion Rates between 10 and 15 Percent



CTBHD = cyclical tree-based hot deck; ML = maximum-likelihood; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.
Note: The ML Methods flowchart is presented in Exhibit 5.4.

Terminal Node 1. The simulation results suggest that WSHD would be adequate as long as the missingness is not dependent on *y* and the X-matrix missingness is less than 2 percent. In addition, WSHD would be a good alternative to discarding a sizeable chunk of the cases, which is what an analyst would be doing if LD or REWT were used. For example, for N4/SPDMON[31] at a deletion rate of 12.5 percent, the X-matrix missingness is about 0.77 percent. The simple hot-deck method used in the simulation was able to correct for much of the bias shown by LD. For relative bias at a deletion rate of 12.5 percent, LD had a median bias measure of 1.8 percent and a mean of 3.2 percent, and WSHD had a median of 1.2 percent and a mean of 1.4 percent (Table F.2).

Table F.6 suggests that a lot of the data being discarded by LD and REWT are useful. For N4/SPDMON, the WSHD, MPLUS, and CTBHD methods all have median ratios of empirical variance to full-sample variance of less than 1.01, suggesting low data loss due to the induction of missingness. Variance underestimation using WSHD was not too severe for the N4/SPDMON model (Table F.10) where the median was -0.7 percent and the mean was -1.6 percent. At worst, variance was underestimated by 10 to 14 percent.

Terminal Node 2. When the X-matrix missingness is greater than or equal to 2 percent, the simulation results suggest that the hot-deck methods underestimate variance more severely. For example, for a deletion rate of 12.5 percent, the N19 models both have X-matrix missingness in the 2.0 to 2.5 percent range (Table 4.20). Even for N19/ABODANL, whose missingness depends on *y* to a lesser degree than for N19/ANLYR, the median relative bias of the variance estimate associated with WSHD is -5.4 percent and the mean is -5.8 percent (Table F.10). The variance underestimation is worse for N19/ANLYR, where the median associated with WSHD is -7.9 percent and the mean is -7.2 percent (Table F.10). The CTBHD method had similar statistics.

If the analyst is willing to accept the variance underestimation and would prefer the improved bias correction of a more complex method (which seems to be better for bias even if the missingness is not highly dependent on *y*), a complex WSHD method is recommended. However, REWT is probably a safer choice. Note that the data loss is not as severe using REWT when X-matrix missingness is higher. For example, for N19/ABODANL in Table F.6, the methods other than LD and REWT show medians and means in the 2 to 4 percent range (as opposed to about 1 percent for N4/SPDMON). So none of the methods that were tested were able to recover all or nearly all of the data lost when missingness was induced, which is as expected because there was still not as much information as there was with full response.

Terminal Node 3. LD and WSHD are not recommended when missingness is dependent on *y* because both show biases that can be addressed easily using REWT or a more complex WSHD method that uses *y* as an auxiliary variable. The best example of an analytic model with

[31] N4/SPDMON is a model of serious psychological distress (SPD) in the past month among women aged 18 to 44. N4/MHTRT is a model of mental health treatment among women aged 18 to 44 with SPD in the past month. N14/YOTMTHLP is a model measuring how much counseling helped adolescents with a major depressive episode (MDE) in the past year who sought counseling in the past year. N14/YORXHLP is a model measuring how much medication helped adolescents with an MDE in the past year who used counseling in the past year. N19/ANLYR is a model of past year pain reliever use among adolescents. N19/ABODANL is a model of past year drug dependence among adolescents with past year pain reliever use.

low X-matrix missingness and missingness dependent on *y* is N4/MHTRT. For this model, the MPLUS method performed very well in the simulation experiment. It was the best at addressing bias (Table F.2), the best at recovering the data loss (Table F.6), and better than even REWT at estimating variance (Table F.10). So, unless the user wants to take a shortcut because the interest is only in statistical significance (see terminal node 5), ML methods are recommended.

Terminal Node 4. REWT is not recommended because too many data are being discarded given the low X-matrix missingness. A more complex hot-deck method than WSHD would hopefully correct for biases due to the missingness being dependent on *y*. Indeed, CTBHD did a noticeably better job than WSHD at correcting for bias for N4/MHTRT (Table F.2).

Terminal Node 5. Analysts reaching this terminal node are unlikely to benefit from a method that recovers the data from the records that were listwise deleted. REWT is recommended to address bias due to the missingness being dependent on *y*.

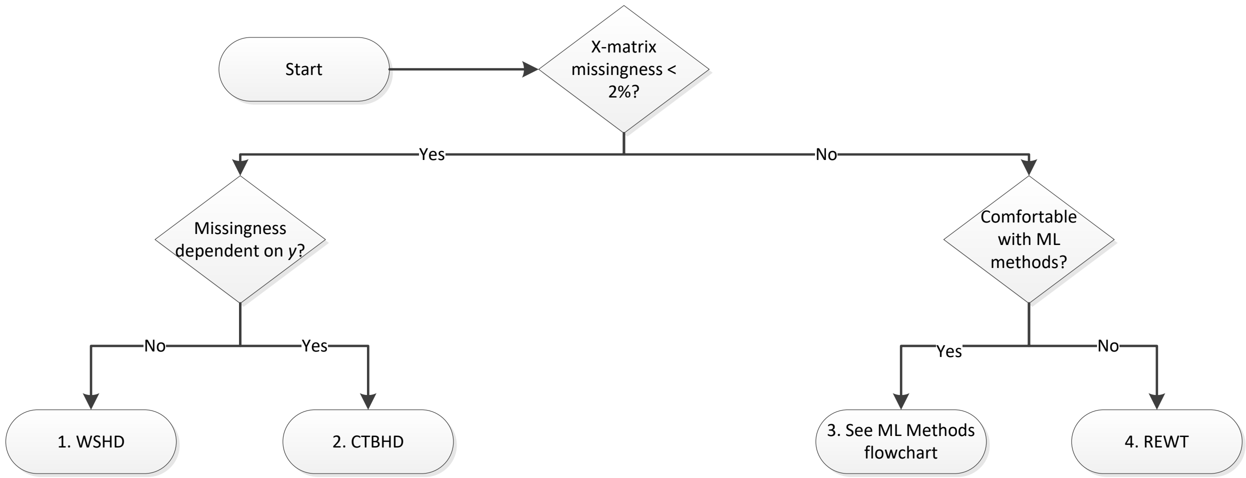
Terminal Node 6. The two N19 analytic models have X-matrix missingness in the 2.0 to 2.5 percent range and missingness dependent on *y* (especially for N19/ANLYR). For both models, REWT did an excellent job correcting for bias (Table F.2). MPLUS did very well for N19/ABODANL but would not run successfully for N19/ANLYR. Compared to REWT with respect to bias correction, the two hot-deck methods performed poorly for both models. Table F.6 suggests that not quite as many data are thrown away for the N19 models using REWT as for the ones where X-matrix missingness is lower. ML methods are recommended if the user is familiar with them and concerned with data loss.

Terminal Node 7. As discussed above for terminal node 6, REWT seems to be a good option for bias correction, and the data loss seems acceptable when the X-matrix missingness is high.

5.3.3 Deletion Rates between 15 and 20 Percent

The decision tree for deletion rates between 15 and 20 percent is presented in Exhibit 5.3. A discussion of the decisions follows the exhibit. Again, when an ML method is selected, it can only be used when the procedure does not fail because the dataset is too large (MPLUS) or when there are at least two PSUs from each stratum in the dataset being analyzed (Latent GOLD). And, as said at the beginning of Section 5.3, the LD method is not recommended for deletion rates in this range due to its limited ability to correct for nonresponse bias.

Exhibit 5.3 Decision Tree for Deletion Rates between 15 and 20 Percent



CTBHD = cyclical tree-based hot deck; ML = maximum-likelihood; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.
Note: The ML Methods flowchart is presented in Exhibit 5.4.

Terminal Node 1. The simulation results suggest that the performance of the hot-deck methods depends on the X-matrix missingness more than anything else. These methods perform well relative to LD and REWT because they do not discard any data, and when the X-matrix missingness is low and the deletion rate is high, LD and REWT are discarding a lot of data. When the missingness is not dependent on *y*, WSHD seems to correct for bias about as well as CTBHD and REWT. Variance underestimation is not as severe when the X-matrix missingness is low. For N4/SPDMON, the best example of low X-matrix missingness and missingness not dependent on *y* (Table 4.10), variance underestimation for WSHD was less than 2 percent for more than 75 percent of the covariates within simulations (Table 4.8).

Terminal Node 2. When the X-matrix missingness is dependent on *y*, CTBHD seems to correct for bias better than WSHD. For N4/MHTRT, the best example of low X-matrix missingness and missingness dependent on *y*, the median bias of CTBHD at a 20 percent deletion rate was 0.6 versus 3.4 percent for WSHD, and the mean was 2.2 versus 5.0 percent for WSHD (Table 4.4). Variance underestimation for CTBHD was tolerable for this model. For most covariates within simulations, it was well below 10 percent (Table 4.8).

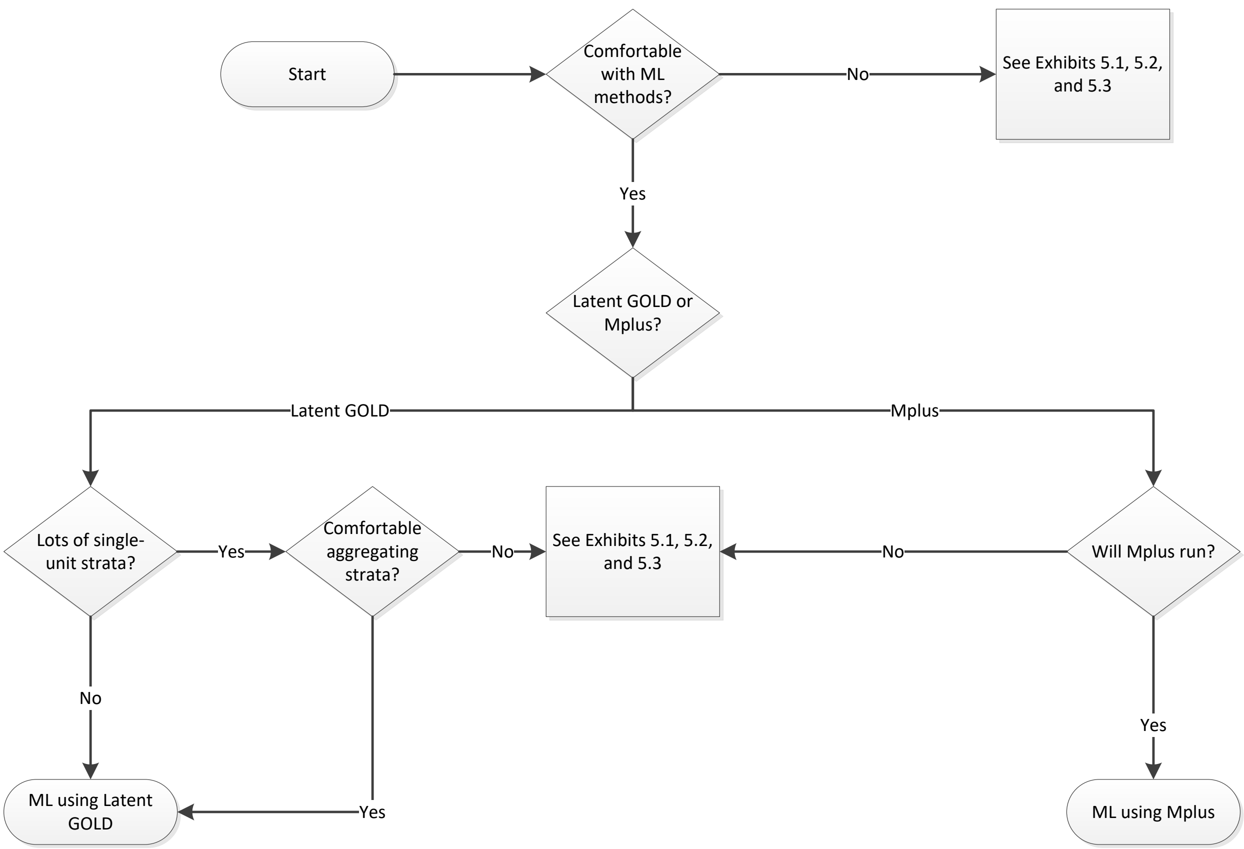
Terminal Node 3. When the X-matrix missingness is greater than 2 percent, more severe underestimation of variance using the hot-deck methods is seen. For the two N19 models, both hot-deck methods have first quartiles well below 10 percent, meaning that for 25 percent of the covariates within simulations, the variance is underestimated by more than 10 percent (Table 4.8). REWT also corrects for bias better than the hot-deck methods for the N19 models (Table 4.4). Theory suggests that REWT should perform better than WSHD because the missingness is dependent on *y*, and it may perform better than CTBHD because of limitations in its modeling approach. If the analyst is familiar with ML methods, they are recommended whether the missingness is dependent on *y* or not so long as there is confidence in the analytic model being fit.

Terminal Node 4. REWT is always a safe option. For both N19 models, REWT performed well for bias correction (Table 4.4) and variance estimation (Table 4.8), and when the X-matrix missingness is high, the data loss is not as severe (Table 4.6).

5.3.4 Decision Tree for ML Methods

The decision trees in Exhibits 5.1, 5.2, and 5.3 refer to an ML Methods flowchart. Exhibit 5.4 presents this flowchart to help users determine whether they should use Latent GOLD software, Mplus software, or an alternative MIVRA method.

Exhibit 5.4 Decision Tree for Analysts Considering ML Methods



ML = maximum-likelihood.

5.4 Conclusions

The main purpose of this report is to aid analysts in determining whether and how they could address missing item values for some of the independent variables of the model in a regression analysis of NSDUH data. Although these results may not apply to datasets derived from other complex sample surveys, the fact that the MIVRA methods that were studied performed as well as they did when applied to real data offers users some assurance in the utility of running extensive simulations on their own datasets. For those who are not able to run their own simulations, caution is recommended when adapting the results of this study to a different survey. In such cases, it would be prudent to apply more than one MIVRA method and to compare the results when the deletion rate is more than 5 percent.

Many users of NSDUH data have access to the WSHD routine in the IMPUTE procedure of SUDAAN, and the CTBHD routine is available by request. Given that these routines are available in software or can be easily recreated, evaluating these two imputation procedures was a focus of this report. Other procedures using donor imputation from mutually exclusive imputation cells, such as the SURVEYIMPUTE procedure in SAS/STAT® 14.1 (SAS Institute Inc., 2015),[32] should produce similar results. It is left to interested readers to confirm this hypothesis.

The two donor imputation methods worked very well, which was better than expected. Even though they failed to capture the added mean squared error due to missing item values being imputed, the missing component of the mean squared error was often not large enough to impact inference. This gave some support to the decision to follow the practice of treating imputed values that were determined using complex (and extensively tested) parametric models as real survey values.

When they can be employed, the ML techniques in Mplus and Latent GOLD have advantages over the other methods investigated here. They use all the data and incorporate the added error due to missing item values and their imputation into the measures of the mean squared error. Mplus appeared to have difficulty with large datasets, and Latent GOLD failed to estimate variances correctly when the dataset was a subsample in which some strata were represented by records from only one PSU. This problem can most easily be dealt with by combining strata into variance strata. If anything, this may lead to a slight overestimation of variances.

As mentioned in Section 3.2.1, LD is not a source of bias when the analytic model being fit holds in the population and the probability of a record being deleted is a function only of the independent variables in the model and not the dependent variable. This holds true even when a record is deleted because the value of its dependent variable is missing. Moreover, the validity of deleting records with missing dependent variables was assumed throughout the evaluation. It was an implied starting point for all the MIVRA methods that were investigated; that is, the assumption was that when dependent variable values were missing, their missingness was wholly a function of the independent variables in the model.

Nevertheless, it is sometimes possible to reweight for nonresponse when a record is listwise deleted because the dependent variable is missing and the deletion is (partly) a function of the dependent variable's value. This can be done by using the WTADJX procedure in SUDAAN and similar procedures in R.[33]

Suppose *S* is a sample to be used to fit a regression model and *D* is the largest subset of *S* containing no model variables, dependent or independent, with missing values. Suppose further

---

[32] Setting METHOD = HOTDECK and SELECTION = WEIGHTED in that procedure parallels setting METHOD = WSHD in SUDAAN's PROC IMPUTE procedure, except that donors are selected with probabilities proportional to the weights with replacement rather than by weighted sequential selection.

[33] One such procedure is "Sampling" (Tille & Matei, 2013).

that the probability that a record k in S is also in D is a known function up to a parameter, $p(\mathbf{z}_k{}^T\boldsymbol{\gamma})$,[34] where $\mathbf{z}_k$ is a vector of model variables, some of which can have missing values. This means that the probability of being listwise deleted is $1-p(\mathbf{z}_k{}^T\boldsymbol{\gamma})$, which is also a function of $\mathbf{z}_k$.

Suppose also that there exists a vector of model variables $\mathbf{q}_k$ with as many components as $\mathbf{z}_k$ that never have missing item values in S (the components can overlap). Under mild conditions, a consistent estimator $\mathbf{g}$ for the parameter γ can be found by solving the calibration equation,

$$\sum_{k\in D}\frac{d_k}{p(\mathbf{z}_k^T\mathbf{g})}\mathbf{q}_k=\sum_{k\in S}d_k\mathbf{q}_k, \qquad 5.1$$

when a solution exists. The sampling weights are represented by d_k in equation 5.1. Their reweighted form is $w_k = d_k / p\left(\mathbf{z}_k{}^T\mathbf{g}\right)$.

There is no test to determine which variables belong in the vector $\mathbf{z}_k$ within the function $p(\mathbf{z}_k{}^T\boldsymbol{\gamma})$ that determines inclusion in D. Instead, this procedure works best in sensitivity analyses. By making alternative assumptions about which variables should be in $\mathbf{z}_k$, including letting $\mathbf{z}_k = \mathbf{q}_k$ as was done in the simulation experiment discussed in Chapter 4, a user can get a feel for how sensitive estimates are to the different assumptions about the missingness process. Users can add assumptions of this sort to a simulation experiment like that described in Chapter 4.

5.5 Promising MIVRA Methods for Future Research

In this section, some MIVRA methods that were not tested in the simulation, but show potential for addressing the issues identified in this study, are discussed.

- *Fractional imputation.* The SURVEYIMPUTE procedure in recently introduced SAS/STAT® 14.1 will perform fully efficient fractional imputation (FEFI; setting METHOD = FEFI). This method, introduced by Kim and Fuller (2004), incorporates the sampling design and creates jackknife replicate weights to estimate variances from its use. Based on its documentation (SAS Institute Inc., 2015), the method appears best suited for datasets in which all the variables are categorical. Unfortunately, Version 14.1 of SAS/STAT® was not available at the time this study was conducted.
- *Listwise deletion with reweighting using Pfeffermann-Sverchkov weights.* Although reweighting often reduced the bias of LD considerably when the dependent variable and other variables in $\mathbf{z}_k$ of $p(\mathbf{z}_k{}^T\boldsymbol{\gamma})$ were never missing (in other words, $\mathbf{z}_k = \mathbf{q}_k$ in equation 5.1), there was usually an increase in the aggregate measure of coefficient mean squared errors from reweighting. This may be because the original complex sampling weights, d_k in equation 5.1, had less variability than the reweighted

[34] WTADJX allows $p(.)$ to have a general form that includes the logistic and truncated logistic function as special cases.

weights, $w_k = d_k / p\left(\mathbf{z}_k{}^T\mathbf{g}\right)$ (where **g** could be estimated in several ways[35]). If the analytic model holds in the population, or comes close enough to holding that one can assume that it does, then one could reweight w_k by running a linear regression of log(w_k) on a function of the independent variables in $\mathbf{z}_k$. Letting t_k be the predicted value of that regression and then dividing each w_k by exp(t_k) produces new revised weights that should be less variable than w_k , yet still removes roughly the same amount of bias in the analytic model being fit as using w_k in place of d_k . This is because, when the regression model holds, multiplying weights that produce nearly unbiased coefficient estimates by any function of the independent variables also produces nearly unbiased coefficient estimates (Pfeffermann & Sverchkov, 1999).[36] Indeed, it is because of this property that using LD does not cause coefficients to be biased when the regression model holds and the probability of inclusion in *D* is wholly a function of the independent variables in the model.

- *Hot-deck imputation that accounts for the variance due to imputation.* Instead of the sorting step of hot-deck imputation, donors can be selected within cells using the approximate Bayesian bootstrap (Rubin & Schenker, 1986) multiple times, and parameter estimation can be conducted using an MI technique. This approach would mitigate the negative bias associated with the variances of the regression coefficients (Section 4.3.2). However, this approach would not provide whatever additional variance reduction may be achieved by sorting. Like WSHD and CTBHD, this methodology assumes that sampled and nonsampled records have the same expectation within cells.
- *Latent GOLD 5.1.* Latent GOLD 5.0 generally performed well in a limited test, and one of the main drawbacks of the method has been addressed in Version 5.1: namely, that the software ignores the variance contribution from single-unit strata. Version 5.1 offers the same options as Stata® for handling strata with only one unit, including the one used by SUDAAN. It is certainly possible that the LG method would perform as well as or better than the MPLUS method if given a full test with Version 5.1.
- *The lavaan.survey package in R.* According to Oberski (2014), this package seems to have similar functionality to Mplus, so future studies may want to investigate whether or not it could be used to implement ML for complex survey data with missing values. One limitation is that the package only supports continuous variables.

[35] As discussed in Chapter 4, *p*(.) was a logistic function, and **g** was computed using a calibration-equation technique rather than weighted logistic regression.

[36] The same semiparametric argument justifies listwise deletion when being deleted is not a function of the dependent variable given the independent variables.

References

Allison, P. D. (2002). *Missing data*. Thousand Oaks, CA: SAGE.

Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference* (pp. 14-16). Retrieved from https://fcsm.sites.usa.gov/reports/research/2005research/

Bennett, D. A. (2001). How can I deal with missing data in my study? *Australian and New Zealand Journal of Public Health, 25*, 64-469.

Bernaards C. A., Belin T. R., & Schafer J. L. (2007). Robustness of a multivariate normal approximation for imputation of incomplete binary data. *Statistics in Medicine, 26,* 1368-1382.

Binder, D. A. (1983). On the variances of asymptotically normal estimators from complex surveys. *International Statistical Review/Revue Internationale de Statistique*, 279-292.

Box, G. E. P., & Draper, N. R., (1987). *Empirical model building and response surfaces*. New York, NY: Wiley.

Center for Behavioral Health Statistics and Quality. (2017a). *2015 National Survey on Drug Use and Health: Methodological resource book (Section 11, Person-level sampling weight calibration)*. Retrieved from https://www.samhsa.gov/data/

Center for Behavioral Health Statistics and Quality. (2017b). *Evaluation of imputation methods for the National Survey on Drug Use and Health*. Retrieved from https://www.samhsa.gov/data/

Cohen, M. P. (2002). Implementing Rao-Shao type variance estimation with replicate weights. *Survey Methodology, 28*, 97-101.

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the American Statistical Association, Survey Research Methods Section, Houston, TX* (pp. 721-726). Washington, DC: American Statistical Association. Retrieved from http://www.amstat.org/ASA/Membership/Sections-and-Interest-Groups.aspx

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., & Presser, S. (2014). Assessing the mechanisms of misreporting to filter questions in surveys. *Public Opinion Quarterly, 78*(3), 721-733.

European Medicines Agency. (2010). *Guideline on missing data in confirmatory clinical trials*. London, UK: Author. Retrieved from http://www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/general/general_content_001221.jsp&mid

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica, 52*, 681-700.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly, 72*, 167-189.

Hallquist, M., & Wiley, J. (2016). *MplusAutomation: Automating Mplus model estimation and interpretation*. R package (Version 0.6-4). Pittsburgh, PA: Author. Retrieved from https://cran.r-project.org/web/packages/MplusAutomation/index.html

Horton, N. J., & Kleinman, K. P. (2007). Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American Statistician, 61*, 79-90.

Jones, M. P. (1996). Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of the American Statistical Association, 91*, 222-230.

Kalton, G., & Kasprzyk, D. (1982). Imputing for missing survey responses. In *Proceedings of the Survey Research Methods Section, American Statistical Association* (pp. 22-31). Washington, DC: American Statistical Association. Retrieved from http://www.amstat.org/ASA/Membership/Sections-and-Interest-Groups.aspx

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*, 184-189.

Kim, J. K., Brick, J. M., Fuller, W. A., & Kalton, G. (2006). On the bias of the multiple-imputation variance estimator in survey sampling. *Journal of the Royal Statistical Society, B 68*, 509-521.

Kim, J. K., & Fuller, W. (2004). Fractional hot deck imputation. *Biometrika, 91*(3), 559-578.

Kim, J. K., & Shao, J. (2013). *Statistical methods for handling incomplete data.* London: Chapman & Hall/CRC.

Kott, P. S. (1995). A paradox of multiple imputation. In *Proceedings of the Survey Research Methods Section, American Statistical Association* (pp. 380-383). Washington, DC: American Statistical Association. Retrieved from http://www.amstat.org/ASA/Membership/Sections-and-Interest-Groups.aspx

Kott, P. S. (2007). Clarifying some issues in the regression analysis of survey data. *Survey Research Methods, 1*, 11-18.

Kott, P. S. (2015). How should I run a regression based on a complex survey that has missing item values? *The Survey Statistician*, 21-26.

Little, R. J. A. (1992). Regression with missing X's: A review. *Journal of the American Statistical Association, 87*, 1227-1237.

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

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data, Second edition*. New York: John Wiley & Sons.

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide, 7th ed*. Los Angeles, CA: Author. Retrieved from http://www.statmodel.com/

National Research Council. (2010). *The prevention and treatment of missing data in clinical trials*. Washington, DC: National Academies Press.

Oberski, D. (2014). lavaan.survey: An R package for complex survey analysis of structural equation models. *Journal of Statistical Software, 57*(1), 1-27. Retrieved from https://www.jstatsoft.org/index

Pfeffermann, D., & Sverchkov, M. (1999). Parametric and semi-parametric estimation of regression models fitted to survey data. *Sankhya, 61*, 166-186.

Pigott, T. D. (2001). A review of methods for missing data. *Educational Research and Evaluation, 7*, 353-383.

R Core Team. (2015). *foreign: Read data stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ... R package* (Version 0.8-66). Retrieved from https://cran.r-project.org/web/packages/foreign/index.html

R Core Team. (2016). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology, 27*, 85-95.

Rao, J. N. K., & Shao, J. (1992). Jackknife variance estimation with survey data under hot deck imputation. *Biometrika, 79*, 811-822.

RTI International. (2013). *SUDAAN® language manual, Release 11.0.1*. Research Triangle Park, NC: Author.

Rubin, D. B, & Schenker, N. (1986). Multiple imputation for interval estimation with ignorable nonresponse. *Journal of the American Statistical Association, 81*, 366-374.

SAS Institute Inc. (2015). *SAS/STAT® 14.1 user's guide*. Cary, NC: SAS Institute Inc.

Satorra, A., & Muthén, B. O. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267-316.

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*, 3-15.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147-177.

Shotwell, M. (2014). *sas7bdat: SAS database reader (experimental)*. R package (Version 0.5). Retrieved from https://cran.r-project.org/web/packages/sas7bdat/index.html

Skinner, C. J. (1989). Domain means, regression and multivariate analysis. In C. J. Skinner, D. Holt, & T. M. F. Smith (Eds.), *Analysis of Complex Surveys* (pp. 59-87). New York: Wiley.

Tille, Y., & Matei, A. (2016). *Package* '*Sampling*'. Survey sampling (Version 2.8). Retrieved from https://cran.r-project.org/web/packages/sampling/index.html

Toth, D., & Eltinge, J. (2011). Building consistent regression trees from complex sample data. *Journal of the American Statistical Association, 106*, 1626-1636.

van Buuren, S., Brand, J. P. L., Groothuis-Oudshoorn, K., & Rubin, D. B. (2006). Fully conditional specification in multivariate imputation. *Journal of Statistical Computation and Simulation, 76*, 1049-1064.

Wilkinson, L., & the Task Force on Statistical Inference, Board of Scientific Affairs, American Psychological Association. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist, 8*, 594-604.

Wine, J., Bryan, M., & Siegel, P. (2014). *2011-12 National Postsecondary Student Aid Study (NPSAS:12) data file documentation* (NCES 2014-182) (pp. 156-157). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/pubsearch/

List of Contributors

This methodological report was prepared by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), and RTI International (a registered trademark and a trade name of Research Triangle Institute). Work by RTI was performed under Contract No. HHSS283201300001C. Peter Tice served as the Government Project Officer, and David Hunter served as the RTI Project Director.

Significant contributors to this report at SAMHSA include Sarra Hedden, Matthew Williams, Jonaki Bose, Rebecca Ahrnsbrak, and Rachel Lipari. Matthew Williams is now with the National Science Foundation.

Significant contributors to this report at RTI include Peter Frechtel, Phillip Kott, Susan Edwards, Rachel Harter, Jeff Laufenberg, Stephen Tueller, and Jean Wang. Other contributors to this report at RTI include Philip Lee, vPaul Biemer, Michael R. Pemberton, Valerie L. Forman-Hoffman, Michael Penne, Christine Davies (formerly with RTI), Lauren Klein Warren, Breda Munoz, Cherie Winder (formerly with RTI), Jeremy Porter, David Heller, Jennifer Cooney, Kevin Wang, and Lance Couzens. At RTI, the report was edited by Claudia M. Clark and formatted by Debbie Bond.

This page intentionally left blank

Appendix A: Details on Reprocessing of Datasets for Detailed Analysis of Missingness

This appendix discusses in detail the reprocessing of the datasets used in the analytic models discussed in Section 2.3.2. The text in this appendix accompanies Tables 2.4 through 2.9.

A.1 Results of Detailed Analysis of Missingness for Study N4

Analytic study N4 involved women aged 18 to 44. Data were pooled across the years 2008 through 2012 to support the analyses.

A.1.1 N4 Model 1

Model 1 in study N4 involved no additional subsetting of the data beyond age and gender. There were no issues with unknown membership in the subpopulation, because the age and gender variables have no missing values. All missing ages are filled in by editing, and the interview cannot continue if the gender is missing, resulting in a unit nonrespondent instead of a unit respondent with item nonresponse. There were 93,121 respondents in the subpopulation.

The dependent variable in the model was the dichotomous version of the Kessler-6 (K6) score, called SPDMON (Kessler et al., 2003), which assesses whether the respondent suffered serious psychological distress (SPD) in the past month. This variable was an adult mental health variable that had the standard zero-fill imputation method undone, as described in Section 2.2.2. After undoing the zero-fill imputation to create an alternate version of SPDMON, 426 respondents in the subpopulation (0.46 percent) had missing values.

Table 2.4 presents missingness statistics for the independent variables. The percentages are based on the 92,695 (93,121 – 426) respondents that were both in the subpopulation and had a nonmissing value for the dependent variable. The revised deletion rate was 4.37 percent, which is 20 percent higher than it was before the zero-fill imputation was undone (3.64 percent). The model included 15 independent variables, 7 of which had no missing values due to editing (age recode) or to the National Survey on Drug Use and Health (NSDUH) complex imputation treatment. The other five variables underwent zero-fill imputation and are described as follows:

- The two drug disorder variables, ABODALC and ABODILL, were substance dependence and abuse variables, for which the zero-fill imputation method was undone, as described in the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* (Center for Behavioral Health Statistics and Quality, 2017b).
- The health problems variable underwent a partial ad hoc zero-fill imputation method for study N4. This variable was derived from 18 dichotomous variables reporting specific health problems in the past year. Respondents reporting zero problems were assigned to the first level, respondents reporting one problem were assigned to the second level, and respondents reporting two or more problems were assigned to the third level. For study N4, respondents with missing values for all 18 source variables were assigned a missing code. For the purposes of this report, the ad hoc zero-fill imputation was completely undone and missing values were inserted where

appropriate. For example, consider a respondent with one negative response and 17 missing values. For study N4, this respondent would be assigned to the first level (no reported problems). For the current NSDUH study, this respondent was assigned a missing code.

- The last two variables, rapid repeat birth (having two biological children within 24 months of age of each other living in the household) and the number of biological children in the household, were created using an ad hoc zero-fill imputation method. The algorithm by which they were created was complex, involving numerous responses to questions in the household roster section of the questionnaire. Given that response rates for these variables tend to be greater than 99 percent, and the effort required to derive alternate versions was high, alternate versions were not created.

A.1.2 N4 Model 2

Model 2 in study N4 focused only on women aged 18 to 44 who reported SPD in the past month according to the SPDMON variable (the same variable that was the dependent variable in N4 model 1). As previously mentioned, 426 respondents (0.46 percent) had a missing value for the subpopulation. There were 7,609 respondents with the alternate version of the SPDMON variable equal to 1.

The dependent variable in this model was the dichotomous variable AMHTXRC3, indicating whether the respondent received mental health treatment in the past year. Of the 7,609 respondents known to be in the subpopulation, 26 (0.34 percent) had missing values for AMHTXRC3.

Table 2.5 presents missingness statistics for the independent variables. The percentages are based on the 7,583 (7,609 – 26) respondents that were both in the subpopulation and had a nonmissing value for the dependent variable. The revised deletion rate was 4.71 percent, which again is much higher than it was before the zero-fill imputation was undone (3.16 percent). There were 17 independent variables in the model: the same 15 used in N4 model 1, plus a lifetime depression indicator (DEPRSLIF2) and a lifetime anxiety indicator (ANXDLIF2). The two new variables did not undergo any imputation, even though they did not have many missing values.

A.2 Results of Detailed Analysis of Missingness for Study N14

Analytic study N14 involved respondents aged 12 to 17 who had a major depressive episode in the past year (YMDEYR = 1). Data were pooled across the years 2006 through 2010 to support the two regression analyses.

The filter questions related to the subpopulation for both models were complex. A detailed subpopulation analysis was undertaken for N14 model 1 to get a sense of the amount of missingness attributable to the filter questions. Many of the same principles apply to N14 model 2. The results of this subpopulation analysis are as follows:

- The subpopulation for N14 model 1 included respondents aged 12 to 17 with YMDEYR = 1 and YOSEEDOC = 1, that is, those who experienced a major depressive episode in the past year and reported seeing some sort of mental health professional about it.

- Of the 111,660 respondents aged 12 to 17 in 2006 through 2010, the subpopulation status was unknown for 303 (0.27 percent). Of these 303 respondents, 173 (57.10 percent) had both YMDEYR and YOSEEDOC missing, 42 (13.86 percent) had YMDEYR = 1 and YOSEEDOC missing, and 88 (29.04 percent) had YMDEYR missing and YOSEEDOC = 1. A further complication is that both subpopulation variables (YMDEYR and YOSEEDOC) were affected by their own set of filter questions.

Given that only 0.27 percent of the respondents had a missing subpopulation indicator for N14 model 1, and there was no reason to expect markedly different results for N14 model 2, it seemed reasonable to continue with only those respondents known to be in the subpopulation for both models. Still, this finding suggests that erosion of item response rates can be due not only to the complexity associated with recodes involving a large number of variables but also to the complexity associated with subpopulation issues.

A.2.1 N14 Model 1

For model 1 in study N14, there were 3,308 respondents known to be in the subpopulation. The dependent variable in the model was YOTMTHLP, a five-level ordinal variable reporting the effectiveness of the treatment or counseling received from the mental health professional. There were 37 respondents in the subpopulation (1.12 percent) that had a missing value.

Table 2.6 presents missingness statistics for the independent variables. The percentages are based on the 3,271 (3,308 – 37) respondents that were both in the subpopulation and had a nonmissing value for the dependent variable. The revised deletion rate was 15.84 percent, compared with the rate of 13.79 percent before zero-fill imputation was undone. There were 14 independent variables in the model. Six of them had no missing values, two did not undergo any imputation, three underwent ad hoc weighted sequential hot-deck (WSHD) imputation within predetermined imputation cells that was undone for the current NSDUH study, and three underwent zero-fill imputation.[37] The three variables that underwent zero-fill imputation are described below:

- The drug disorder variable was a substance dependence and abuse variable, which had the zero-fill imputation method undone as described in Section 2.2.2.
- The two count variables (number of delinquent behaviors and family encouragement) were similar in nature to the health problems variable previously described for study N4. The versions used in the N14 analyses were based on the number of affirmative responses to a set of yes-no questions, without regard to missingness. The alternate versions used in the current NSDUH study have missing values where appropriate.

[37] Of the five variables that did not undergo NSDUH complex imputation, the three with the most missingness underwent ad hoc WSHD to support the analyses in the original N14 manuscript.

As shown in Table 2.6, the two variables with the most missingness are grades and number of mental health visits.

A.2.2 N14 Model 2

For model 2 in study N14, the subpopulation included respondents aged 12 to 17 who had a major depressive episode in the past year and reported taking medication for mental health reasons in the past year. There were 221 cases (0.20 percent) with a missing value for the subpopulation indicator and 1,545 respondents known to be in the subpopulation.

The dependent variable in this model was YORXHLP, a five-level ordinal variable reporting the effectiveness of the medication(s). Of the 1,545 respondents known to be in the subpopulation, 6 (0.39 percent) had a missing value.

Table 2.7 presents missingness statistics for the independent variables. The percentages are based on the 1,539 (1,545 – 6) respondents that were both in the subpopulation and had a nonmissing value for the dependent variable. The revised deletion rate was 17.15 percent, compared with the rate of 15.14 percent before zero-fill imputation was undone. There were 13 independent variables in the model: the same 14 that were used in N14 model 1, minus the one for past year mental health medications, which all subpopulation members reported taking. For N14 model 2, the two variables with the most missingness are grades and number of mental health visits (the same as for N14 model 1).

A.3 Results of Detailed Analysis of Missingness for Study N19

Analytic study N19 involved nonmedical use of pain relievers among respondents aged 12 to 17. Data were pooled across the years 2008 through 2012 to support the analyses. Although four models were involved in this study (Table 2.1), only two are considered here, because three of the models differ only in their use of interaction terms. Interaction terms do not affect the deletion rate. Therefore, there was little benefit from considering these three similar models separately.

A.3.1 N19 Model 1

The subpopulation for model 1 in study N19 included all respondents aged 12 to 17. There were 112,519 respondents in the subpopulation. The dependent variable in the model was ANLYR, a yes-no variable indicating past year nonmedical use of pain relievers. This variable had no missing item values because it underwent the NSDUH complex imputation treatment.

Table 2.8 presents missingness statistics for the independent variables. The revised deletion rate was 13.14 percent, compared with the rate of 10.33 percent before zero-fill imputation was undone. There were 12 independent variables in the model. Five of them had no missing values, three underwent ad hoc WSHD imputation within predetermined imputation cells that was undone for the current NSDUH study, and four underwent zero-fill imputation.[38] The four variables that underwent zero-fill imputation are described below:

- The two drug disorder variables were substance dependence and abuse variables, which had the zero-fill imputation method undone as described in Section 2.2.2.
- The two count variables (family support and number of delinquent behaviors) were similar in nature to the health problems variable previously described for study N4 and to the count variables previously described for study N14. The versions used in the N19 analyses were based on the number of affirmative responses to a set of yes-no questions, without regard to missingness. The alternate versions used in the current NSDUH study have missing values where appropriate.

As shown in Table 2.8, the two variables with the most missingness are family support and grades.

A.3.2 N19 Model 2

For model 2 in study N19, the subpopulation included respondents aged 12 to 17 who used pain relievers in the past year without a prescription. The variable used to create the subpopulation was ANLYR, which was the dependent variable in N19 model 1. As previously discussed, this variable underwent NSDUH complex imputation treatment and had no missing values. Therefore, there were no respondents whose subpopulation status was unknown. There were 7,084 cases in the subpopulation.

The dependent variable in this model was pain reliever use disorder in the past year, a binary variable with responses of yes or no. This was one of the variables that underwent zero-fill imputation that was undone (Center for Behavioral Health Statistics and Quality, 2017b). Of the 7,084 respondents in the subpopulation, 788 (11.12 percent) had a missing value. The relatively high percentage of respondents with missing values for the dependent variable presents a cause for concern.[39] However, for the purposes of this report, cases with a missing value for the dependent variable are dropped from the dataset.

Table 2.9 presents missingness statistics for the independent variables. The percentages are based on the 6,296 (7,084 – 788) respondents that were both in the subpopulation and had a nonmissing value for the dependent variable. The revised deletion rate was 12.58 percent, compared with the rate of 10.28 percent before zero-fill imputation was undone. The same 12 variables used in N19 model 1 were used in N19 model 2. The missingness rate may be slightly lower for this model due to the elimination of cases with a missing dependent variable.

---

[38] As for N14, the three variables that did not undergo NSDUH complex imputation underwent ad hoc WSHD to support the analyses in the original N19 manuscript.

[39] See Chapter 5 for a brief discussion of ways to deal with missingness in the dependent variable of the model.

This page intentionally left blank

Appendix B: Missingness Statistics for 55 NSDUH Analytic Models

This appendix includes a more detailed version of Table 2.1. Missingness statistics are provided for each of the 55 models involved in the preliminary screening (Section 2.2.1).

Table B.1 Missingness Statistics for 55 Models across 16 NSDUH Studies

Study[1]	Model	Number in Domain	Percentage Dropped: Due to Missing Y	Percentage Dropped: Due to Missing X[2]	Prevalence of Y: Cases with Nonmissing Y	Prevalence of Y: Cases with Nothing Missing
C8	1	68,900	0.01	2.38	2.53	2.53
	2	17,600	0.01	3.22	5.36	5.36
C10	1	229,600	0.01	6.05	5.75	5.66
	2	229,600	0.01	5.51	5.75	5.68
	3	229,600	0.01	5.51	5.75	5.68
I1	1	113,800	0.00	14.52	20.86	21.12
	2	121,800	0.00	14.26	6.46	6.82
	3	159,000	0.00	13.33	11.01	11.44
	4	159,000	0.00	13.33	5.87	6.09
	5	62,400	0.00	2.59	7.71	7.75
	6	181,500	0.00	2.30	38.17	38.15
	7	181,500	0.00	2.30	31.33	31.32
I2	1	78,000	0.00	0.60	3.48	3.46
P4	1	52,300	0.00	0.00	No differences; all model variables underwent imputation	
	2	107,900	0.00	0.00		
	3	31,100	0.00	0.00		
	4	48,300	0.00	0.00		
T1	1	180,000	0.00	13.83	9.67	10.06
	2	180,000	0.00	13.83	3.40	3.51
	3	18,200	0.01	10.89	Continuous response variable	
T2	1	319,320	0.00	2.59	25.37	25.36
	2	319,320	0.00	2.59	18.45	18.48
	3	102,291	0.00	2.76	Continuous response variable	
N1	1	21,300	0.00	1.31	8.92	8.90
	2	21,300	0.00	1.31	90.40	90.41
	3	21,300	0.00	1.31	13.07	13.13
	4	3,000	0.01	1.11	44.45	44.68
N4	1	93,100	0.00	3.64	7.10	7.11
	2	7,600	0.00	3.16	51.00	51.19
N14	1	3,300	1.12	13.79	{17.84, 21.98, 24.36, 24.39, 11.42}[3]	{17.95, 21.91, 23.94, 24.18, 12.02}[3]
	2	1,500	0.39	15.14	{16.49, 17.28, 21.90, 28.89, 15.44}[3]	{16.37, 16.58, 21.61, 29.84, 15.60}[3]
N15	1	70,700	0.00	0.00	Negligible differences	
	2	92,100	0.00	0.00		
	3	82,900	0.00	0.00		
N18	1	22,600	0.43	0.86	20.40	20.52
	2	22,600	0.07	0.86	3.55	3.54
	3	22,600	0.11	0.86	25.40	25.49
N19	1,3,4	112,500	0.00	10.33	6.10	6.30
	2	7,100	0.00	10.28	14.60	15.30

Table B.1 Missingness Statistics for 55 Models across 16 NSDUH Studies (continued)

Study[1]	Model	Number in Domain	Percentage Dropped		Prevalence of Y	
			Due to Missing Y	Due to Missing X[2]	Cases with Nonmissing Y	Cases with Nothing Missing
PR2	1	21,500	0.00	25.56	13.20	11.30
	2	14,100	0.00	26.93	14.40	12.60
PR5a	1	500,200	0.00	0.01	14.27	14.27
	2	500,200	0.00	0.01	10.59	10.59
	3	500,200	0.00	0.01	6.10	6.10
	4	500,200	0.00	0.01	4.60	4.60
	5	500,200	0.00	0.01	1.53	1.53
	6	500,200	0.00	0.01	0.56	0.56
	7	500,200	0.00	0.01	7.62	7.62
	8	365,200	0.01	0.01	6.64	6.64
	9	229,600	0.00	0.01	18.06	18.06
	10	229,600	0.00	0.01	3.88	3.88
	11	229,600	0.00	0.01	3.76	3.76
PR7	1	420,000	0.00	4.77	2.60	2.72

[1] These 16 studies are identified by the NSDUH analytic study codes used internally by SAMHSA.
[2] The percentage of observations with nonmissing Y have no missing X.
[3] For study N14, Y has five levels. The prevalence is the (weighted) percentage of cases in each of the five levels.

Appendix C: How to Implement Listwise Deletion with Reweighting and Weighted Sequential Hot-Deck Imputation

C.1 Introduction

This appendix describes the listwise deletion with reweighting (REWT) and weighted sequential hot-deck (WSHD) methods used for handling missing item values in regression analyses. These methods are discussed in the body of the report and are used in the simulation study described in Chapter 4. Sample code is provided in SAS®,[40] SAS®-callable SUDAAN®,[41] and Stata®. The WSHD method is not available in Stata. Therefore, the sample code is for a simpler unweighted hot-deck imputation.

These two methods are described in the same appendix because both involve two steps—treating the data and then running the analytic model on the treated data—and they differ only in the treatment. For REWT, the treatment involves adjusting the weights. For WSHD, the treatment involves replacing missing values in the independent variables with valid values.

The dataset used in the SAS and SAS-callable SUDAAN sample code is assumed to have variables associated with the analytic model, named as follows:

- *depVar*, the dichotomous dependent variable;
- a set of never-missing independent variables;
- a set of sometimes-missing (incomplete) independent variables; and
- *subpop*, a 0/1 indicator of membership in the population of interest.

The rest of the variables have the names used on typical National Survey on Drug Use and Health (NSDUH) datasets:

- *analwt*, the analysis weight;
- *vestr*, the variance estimation stratum; and
- *verep*, the variance estimation replicate/primary sampling unit;
- *irsex,* the imputation-revised gender;
- *race4,* the imputation-revised, four-level race/ethnicity variable; and
- *age,* the respondent's age, which is never missing.

C.2 Reweighting

The REWT method involves (1) reweighting the complete records to represent themselves and the listwise-deleted records, and (2) fitting the original analytic model using the complete observations and their adjusted weights.

For the simulation study described in Chapter 4, the reweighting step was completed using the WTADJUST procedure in SAS-callable SUDAAN. This procedure was used in part because it reduces the number of steps required. Users unfamiliar with this procedure or without

[40] SAS® software is a registered trademark of SAS Institute Inc.
[41] SUDAAN® is a registered trademark of Research Triangle Institute.

access to SUDAAN (or to an analogous procedure in R) will get similar results using dichotomous logistic regression to estimate the probability that a record is complete.[42] The steps required are the following:

1. Create a 0/1 indicator of missingness in the covariates, called *completeX*. This will be equal to 1 if the record has no missing independent variables, equal to 0 if the record has one or more missing independent variables, and missing if either the record has a missing value for the dependent variable or the record is not a member of the subpopulation of interest.
2. Use the WTADJUST procedure to create the adjusted weight, called *wtfinal*, by modeling the missingness indicator as a function of the dependent variable from the analytic model and the never-missing independent variables from the analytic model.

Sample code in SAS for the first step is below.

```
%let indepVarsCOMPLETE=<complete independent variables>;
%let indepVarsINCOMPL=<incomplete independent variables>;
data withMissInd;
  set <original dataset>;
  if subpop=1 and depVar ne . then do;
    if nmiss(&indepVarsINCOMPL)=0 then completeX=1;
    else completeX=0;
  end;
  nqid=questid+0;
run;
```

Sample code in Stata for the first step is below.

```
use <original dataset>
gen completeX=0 if missing(<incomplete independent variables>)
      & !missing(depVar) & subpop==1
replace completeX=1 if missing(complete) & !missing(depVar)
      &  subpop==1
```

Sample code in SAS-callable SUDAAN for the second step is below. The output dataset includes the adjusted weight, called *wtfinal*. The complete cases, which have *completeX* = 1, have *wtfinal* ≥ *analwt*. They are reweighted to account not only for the nonsampled cases (which is what *weight analwt* does) but also for incomplete cases (i.e., cases with *completeX* = 0). The code below assumes that all the never-missing independent variables are categorical, and thus they belong on the CLASS statement.

[42] See Section 24.2.2 of RTI International (2013) for a description of the model used by the WTADJUST procedure, and see Section 24.2.3 for a comparison of the method used by the WTADJUST procedure with the method involving dichotomous logistic regression.

```
proc wtadjust data=withMissInd adjust=nonresponse design=wr;
  nest vestr verep;
  weight analwt;
  subpopx subpop=1 & depVar ne .;
  class depVar &indepVarsCOMPLETE;
  model completeX=depVar &indepVarsCOMPLETE;
  idvar vestr verep subpop depVar &indepVarsCOMPLETE
                                          indepVarsINCOMPL;
  output idvar wtfinal / filename=withAdjWt replace;
run;
```

As mentioned above, an alternate approach that does not use the WTADJUST procedure in SUDAAN and would produce similar adjusted weights involves the following steps:

1. Fit a dichotomous logistic regression model with *completeX* as the dependent variable and *depVar* and the complete independent variables from the analytic model as the independent variables. Save the predicted values from the model, which are the estimated probabilities that each case is complete in the independent variables given that it is both in the subpopulation of interest and has a nonmissing value for the dependent variable.
2. Create a nonresponse adjustment weight factor that is equal to 0 when *completeX* is 0 and equal to the reciprocal of the estimated probabilities from step 1 when *completeX* is 1.
3. Multiply the analysis weight *analwt* by the nonresponse adjustment factor from step 2 to get the adjusted weight *wtfinal*.

These steps can be performed in any statistical software. Sample code in Stata for this step is below.

```
logit completeX <complete independent variables>
predict predVal
generate nrAdj=1/predVal
replace nrAdj=0 if completeX==0
generate wtfinal=analwt*nrAdj
```

C.3 WSHD

The WSHD method involves two steps: (1) implementing multivariate imputation using WSHD for all sometimes-missing independent variables in the analytic model, and (2) fitting the analytic model using the imputed values from step 1. WSHD is described in detail in Cox (1980) and RTI International (2013).

For the simulation study described in Chapter 4, the WSHD imputation step was completed using the IMPUTE procedure in SAS-callable SUDAAN. Imputation classes were formed using *irsex* and *race4*, but any variables deemed reasonable by the analyst can be used.[43]

[43] The classing (and sorting) variables might be selected by an analyst based on expert judgment, exploratory data analysis, modeling using regression and classification trees or parametric regression, or any combination of these.

However, using more classing variables and/or classing variables with rare levels increases the probability of classes with no item respondents, which causes the procedure to fail. Within the imputation classes, records were sorted by *age* and then *analwt*. The use of sorting variables allows analysts to exploit information in continuous auxiliary variables as well as in extra categorical auxiliary variables that might cause small classes if they were used as classing variables. The output dataset, called *imputed*, will have no missing values in the previously incomplete independent variables.

```
%let indepVarsCOMPLETE=<complete independent variables>;
%let indepVarsINCOMPL=<incomplete independent variables>;
proc impute data=<original data set> method=WSHD seed=12345 notsorted;
  weight analwt;
  subpopx subpop=1 & depVar ne .;
  class &indepVarsINCOMPL;
  impby IRSEX RACE4 ;
  ICSORT AGE ANALWT;
  impvar &indepVarsINCOMPL;
  idvar vestr verep subpop depVar &indepVarsCOMPLETE;
  impname / overwrite;
  output /impute=default filename=imputed replace;
run;
```

The WSHD method is not widely available in other statistical software packages. Sample code for a comparable (but unweighted) hot-deck procedure in Stata is below. This code uses the *hotdeck* program available from the SSC library (command: .ssc install hotdeck). After running *hotdeck*, the imputed values are stored in the dataset *imp1* and then merged back onto the original dataset, replacing only those values that were missing.

```
use <original data set>
hotdeck <incomplete independent variables> using imp, store
  by(irsex race4) seed(12345) keep(vestr verep questid) impute(1)
clear
use imp1
drop if completeX==1 //keep only imputed records
save imp1
clear
use <original data set>
merge 1:1 vestr verep questid using imp1, update
```

C.4 Fitting the Analytic Model

The code used to fit the analytic model is similar whether REWT or WSHD was used to treat the data. If REWT was used, then the adjusted weight *wtfinal* should be used to fit the model. If WSHD was used, then the imputation-revised dataset should be used along with the original analysis weight *analwt*. Sample code in SAS-callable SUDAAN for this step, assuming REWT was used, is below. The code assumes that all the never-missing independent variables and all the sometimes-missing independent variables are categorical, and thus they belong on the

CLASS statement in SUDAAN. The code also assumes that the analyst is interested in fitting a dichotomous logistic regression model. Other analyses would require slightly modified code, but the idea is the same.

```
proc rlogist data=withAdjWt design=wr;
  nest vestr verep;
  weight wtfinal;
  subpopx subpop=1;
  class &indepVarsCOMPLETE &indepVarsINCOMPL;
  model depVar=&indepVarsCOMPLETE &indepVarsINCOMPL;
  print beta sebeta t_beta p_beta waldf waldp / style=nchs;
run;
```

Sample code to fit the analytic model in Stata, assuming REWT was used, is below.

```
svyset verep [pw=wtfinal], strata(vestr) singleunit(centered)

svy: glm depVar <complete and incomplete independent variables>, ///
  family(binomial)
```

This page intentionally left blank

Appendix D: How to Implement Full Information Maximum Likelihood Logistic Regression in Mplus®

D.1 Introduction

This appendix describes the full information maximum likelihood (FIML) approach using complex survey data in the Mplus®[44] software package (Muthén & Muthén, 1998-2015). Included is a demonstration for complete case analysis, showing that the SUDAAN®[45] and Mplus software programs give identical results.

The ordered categorical (ordinal) outcome variable YOTMTHLP is considered in the National Survey on Drug Use and Health (NSDUH) data presented in this appendix. This variable is on a 5-point Likert-type scale indicating whether counseling helped the respondent (i.e., not at all, a little, some, a lot, or extremely). This variable is modeled using ordered multinomial logistic regression. The independent variables include survey year, age, gender, race/ethnicity, family income, health insurance status, rural/urban domicile, number of past year delinquent behaviors, grades in school, parental encouragement, number of religious services attended in the past year, severe role impairment, number of mental health visits in the past year, and the status of a prescription for mood medication in the past year.

D.2 FIML in Mplus

Mplus is syntax driven, though the initial syntax for (ordered) logistic regression can be initiated using the Mplus language generator. Following are characteristics of Mplus and its syntax language:

- Variable names in Mplus must be eight characters or fewer, must start with a letter, and can contain numbers and the underscore character ("_").
- Lines can be no longer than 90 characters, including spaces.
- Mplus has no analog to the SAS®[46] and SUDAAN class statements, which requires the user to appropriately dummy code categorical independent variables. This is not required of dependent variables, which can be declared as binary or ordinal using the CATEGORICAL command[47] or nominal using the NOMINAL command. As shown later in this section, the DEFINE command is an internal Mplus facility to create the dummy codes in Mplus, although this can be done in general purpose software prior to exporting the data for Mplus.
- The comment character is "!".
- Commands end with a colon (":").
- Subcommands, transformations, and model statements end with a semicolon (";").
- In subcommands, the keywords "=", "are", and "is" can be used interchangeably.

---

[44] Mplus® is a registered trademark of Muthén & Muthén.
[45] SUDAAN® is a registered trademark of Research Triangle Institute.
[46] SAS® software is a registered trademark of SAS Institute Inc.
[47] Mplus automatically detects the number of levels in the dependent variable and is always fitting an ordered logistic regression model, even if the outcome is binary, because binary logistic regression is a special case of ordered logistic regression.

- Mplus can read in plain text tab delimited (e.g., "*.dat") or comma separated (*.csv) files (more detail is provided later in this section).
- The *Mplus User's Guide* (Muthén & Muthén, 1998-2015) can be downloaded from http://statmodel.com/ugexcerpts.shtml.
- The Mplus technical appendices to the user's guide are available at http://statmodel.com/techappen.shtml.
- The Mplus discussion board can be searched at http://statmodel.com/cgi-bin/discus/discus.cgi or by doing a Google search for "search terms here site:statmodel.com".

An Mplus input file begins with a TITLE statement. Any user notes can be placed here. There is no limit to the number of lines in the title statement, and there are no syntax constraints within this section. The notes will be printed in the output file. As shown in Exhibit D.1, the name of the input file is placed in the title.

The second section is the DATA statement. This gives the location of the data. If the data are in the same directory as the input file, only the file name for the data needs to be supplied; otherwise, the full path must be given. Data must be plain text ASCII files with no row or column (variable) names. Only numeric data are allowed in the data file. Format options are described in detail in the *Mplus User's Guide* (Muthén & Muthén, 1998-2015). By default, Mplus assumes data are in free format, with missing data designated by a number (e.g., -999), the "." character, or the "*" character. The "." character is the default in SAS (e.g., when exporting to a csv file, which Mplus can read as long as the above requirements are met).

R users (R Core Team, 2016) may consider using the package MplusAutomation (Hallquist & Wiley, 2016). R can read uncompressed sas7bdat files using the R package sas7bdat (Shotwell, 2014) or using the read.ssd() function in the foreign package (R Core Team, 2015), which writes and executes a SAS script to export the SAS dataset and read it into R. In addition, csv files can be exported from SAS and then read into R with the native read.csv() function (in which case missing values can be read in as blanks into R, and MplusAutomation will output them as "."). The PrepareMplusData() function of the MplusAutomation package can then be used to force data to comply to Mplus data requirements and will write a skeleton Mplus input file with populated DATA and VARIABLE commands. Alternatively, Mplus can read csv files directly if there are no variable names in the first row, no character values, and no values are in quotation marks.

The DEFINE command allows users to recode data. Exhibit D.1 shows each independent variable being dummy coded. When making a comparison, the terms EQ (equals), NE (not equal to), GT (greater than), GE (greater than or equal to), LT (less than), or LE (less than or equal to) are required, and assignment of a value is done with the "=" operator. Other options and available transformations are provided in the *Mplus User's Guide* (Muthén & Muthén, 1998-2015). In the model commands (described below), the user selects the reference category by excluding that category's dummy variable from the analysis. At the end of the DEFINE command in the lower panel of Exhibit D.1, the dependent variable "yotmthlp" is reverse coded. For logistic regression, Mplus estimates thresholds (which are re-parameterizations of the intercepts) instead of intercepts. For illustration purposes, reverse coding the dependent variable ensures regression coefficients in the Mplus output have the same sign as the SUDAAN output.

The VARIABLE command has several subcommands. The NAMES subcommand lists all the variable names in the dataset. If the number of names is greater than the number of variables in the dataset, an error will occur (the complement does not hold; fewer names than the number of variables in the dataset will not lead to an error). It is left to the user to ensure that the order of the variables in the names statement matches the order of the variables in the dataset. As noted above, variable names can be only eight or fewer characters in length. The MISSING subcommand designates the missing value flag, which can be one or more numbers (e.g., -999) or the "." or "*" symbols (but not both, and not a combination of numbers or either of those two characters). Additional options for the missing data flag are provided in the *Mplus User's Guide* (Muthén & Muthén, 1998-2015).

The WEIGHT, STRATIFICATION, CLUSTER, and SUBPOPULATION[48] subcommands allow users to specify weight, stratum, and primary sampling unit variables, respectively, and subpopulation restrictions. At least one of the first three is required to invoke the Mplus complex survey data procedure and must be used with the TYPE IS COMPLEX subcommand of the ANALYSIS command, as described later.

The IDVAR subcommand allows for an identification variable to be specified. The presence of this option only affects whether the identification variable is saved to any output datasets (see the SAVEDATA command in the *Mplus User's Guide* in Muthén & Muthén [1998-2015]). The USEVARIABLES subcommand specifies which subset of variables in the NAMES subcommand of the VARIABLE command and the DEFINE command are to be used in the analysis. If this subcommand is omitted, Mplus will include all variables in the analysis even if they do not appear in the model command. Variables created in the DEFINE command must follow variables from the NAMES subcommand.

The CATEGORICAL subcommand specifies which dependent variables have a binary or ordinal distribution. Independent variables cannot be included in the CATEGORICAL subcommand. Additional options that are not described herein are subcommands for nominal (i.e., to obtain unordered multinomial logistic regression), count (Poisson, zero inflated Poisson, negative binomial, zero inflated negative binomial, negative binomial hurdle), survival, and censored (e.g., Tobit regression) dependent variables.

The ANALYSIS command controls the model estimation options. The TYPE subcommand specifies the type of analysis, and Exhibit D.1 shows TYPE IS COMPLEX, which, in conjunction with the WEIGHT, STRATIFICATION, SUBPOPULATION, and/or CLUSTER subcommands of the VARIABLE command, invokes the complex survey pseudo-maximum likelihood (PML, a form of FIML; Skinner, 1989) estimation procedure of Mplus (Asparouhov & Muthén, 2005). The TYPE IS COMPLEX command must be used when WEIGHT, STRATIFICATION, or SUBPOPULATION commands are used. It is not required for the CLUSTER option (which is also available for multilevel models under TYPE IS TWOLEVEL).

[48] The subpopulation option is available for most analyses, but not all, in Mplus. If using complete case analyses where the subpopulation option is not available, users should create a copy of the dependent variable with values set to missing for cases not in the subpopulation of interest. To see that this is the case, the user can take an analysis that does work with the subpopulation command and fit it in either SUDAAN or Mplus with the original dependent variable and the subpopulation command, and then a second time using the altered dependent variable and excluding the subpopulation command.

The estimator is specified here as a maximum likelihood estimator with robust standard errors (MLR), which applies a PML estimator with robust standard errors. This estimator is required to replicate results in SUDAAN. Alternatives include various weighted least squares estimators.[49]

The MODEL statement allows one or more equations to be specified. Multiple equation systems lead to seemingly unrelated regression, mediation, and structural equation models. The dependent variable precedes the keyword ON. The independent variables appear on the right-hand side of the ON keyword. This can be specified equivalently by a separate statement for each independent variable, as shown in the following example:

```
YOTMTHLP ON y2006;
YOTMTHLP ON y2007;
…
YOTMTHLP ON yesmeds;
```

This alternative is important when the user wants to impose linear or nonlinear constraints on model parameters (see the MODEL CONSTRAINT command in the *Mplus User's Guide* in Muthén & Muthén [1998-2015]).

The last command is the OUTPUT command where standardized coefficients are requested using the keyword STDYX. The keyword TECH1 shows all the parameter matrices available to a given set of variables under the given model type, placing zeros where parameters are not estimated (i.e., fixed to zero or some other constant) and placing numbers where parameters are estimated. This is useful for checking the specification of complex, multi-equation models. The order of the numbers indicating estimated parameters is useful for reading additional types of output available in the SAVEDATA command, which is not described herein.

D.3 Complete Case Analyses in Mplus and SUDAAN

The complete case analyses in this section use both SUDAAN and Mplus to show that SUDAAN results can be replicated in Mplus. The Binder (1983) method of standard error estimation is used in Mplus (Satorra & Muthén, 1995) and is the default option in SUDAAN (RTI International, 2013).

Exhibit D.1 shows the ordered multinomial logistic regression syntax for SUDAAN and Mplus. In SUDAAN, sampling with replacement (DESIGN = WR) and an independent correlation structure (R = INDEPENDENT) is specified. The stratification and cluster variables are specified in SUDAAN using the nest statement (nest vestr verep;) and in Mplus using separate stratification (STRATIFICATION = vestr;) and cluster (CLUSTER = verep;) subcommands. The weight variable is specified similarly in SUDAAN (weight analwt;) and Mplus (WEIGHT = analwt;). In SUDAAN, categorical dependent and independent variables are shown in the class statement. The model statement ends with the CUMLOGIT options specifying cumulative (ordered) multinomial logistic regression. This is achieved in Mplus by specifying the dependent variable on the CATEGORICAL subcommand of the VARIABLES command. In

[49] Bayesian estimation with complex survey data has not yet been implemented in Mplus. Requesting Bayesian estimation in Mplus with TYPE IS COMPLEX will revert to the default estimator WLSMV. The WLSMV estimator will not replicate SUDAAN results.

SUDAAN, the "SETENV DECWIDTH=3;" option is specified to obtain results to three decimal places. In Mplus, the number of decimals printed to the output file is three and cannot be changed.[50] The variables delinquency, encouragement, and religServices are shortened to delinq, encourage, and religsrv in Mplus to comply with the eight character length constraint. Table D.1 shows that results of the two analyses are identical.

Exhibit D.1 SUDAAN® and Mplus® Syntax for Complete Case Ordered Multinomial Logistic Regression Analysis

```
SUDAAN Syntax
proc multilog data = data.mhtrt0610 filetype = sas notsorted
DESIGN=WR SEMETHOD=BINDER R=INDEPENDENT MAXITER=20;

      nest vestr verep;
      weight analwt;
      subpopn yotmthlp ^= .;

      class year yotmthlp age irsex race4 income5 irinsur4
            rururb00 delinquency grades encouragement
            religServices mdeimpy smhvst meds;

      model yotmthlp = year age irsex race4 income5 irinsur4
            rururb00 delinquency grades encouragement religServices
            mdeimpy smhvst meds / cumlogit;

      SETENV DECWIDTH=3;
run;

Mplus Syntax
TITLE: "logistic regression - defaults.inp"
DATA: FILE = "S:/RTI Shares/NSDUH/FIMLguidance/Mplus/Data/mhrtr0610.dat";

! Use the define statement to create dummy variables (Mplus does not have a
! SAS/SUDAAN analog of a class statement); alternatively, this can be done
in
! another program such as SAS prior to exporting the data for Mplus
DEFINE:
y2006 = 0;
y2007 = 0;
y2008 = 0;
y2009 = 0;
y2010 = 0;
if year eq 2006 then y2006 = 1;
if year eq 2007 then y2007 = 1;
if year eq 2008 then y2008 = 1;
if year eq 2009 then y2009 = 1;
if year eq 2010 then y2010 = 1;
```

[50] The SAVEDATA command can be used to obtain parameter estimates, standard errors, and the information matrix with precision up to eight decimal places in separate free format data files, requiring additional row and column labeling to be read easily. See the RESULTS and TECH3 subcommands of the SAVEDATA command in the *Mplus User's Guide* (Muthén & Muthén, 1998-2015).

Exhibit D.1 SUDAAN® and Mplus® Syntax for Complete Case Ordered Multinomial Logistic Regression Analysis (continued)

```
age12 = 0;
age13 = 0;
age14 = 0;
age15 = 0;
age16 = 0;
age17 = 0;
if age eq 12 then age12 = 1;
if age eq 13 then age13 = 1;
if age eq 14 then age14 = 1;
if age eq 15 then age15 = 1;
if age eq 16 then age16 = 1;
if age eq 17 then age17 = 1;

male   = 0;
female = 0;
if irsex eq 1 then male   = 1;
if irsex eq 2 then female = 1;

white = 0;
black = 0;
other = 0;
hisp  = 0;
if race4 eq 1 then white = 1;
if race4 eq 2 then black = 1;
if race4 eq 3 then other = 1;
if race4 eq 4 then hisp  = 1;

inlt20k  = 0;
in2050k  = 0;
in5075k  = 0;
in75100k = 0;
ingt100k = 0;
if income5 eq 1 then inlt20k  = 1;
if income5 eq 2 then in2050k  = 1;
if income5 eq 3 then in5075k  = 1;
if income5 eq 4 then in75100k = 1;
if income5 eq 5 then ingt100k = 1;

insure = 0;
uninsure = 0;
if irinsur4 eq 1 then insure = 1;
if irinsur4 eq 2 then uninsure = 1;

rural = 0;
urban = 0;
if rururb00 eq 1 then rural = 1;
if rururb00 eq 2 then urban = 1;
```

Exhibit D.1 SUDAAN® and Mplus® Syntax for Complete Case Ordered Multinomial Logistic Regression Analysis (continued)

```
delinq0 = 0;
delinq1 = 0;
delinq2 = 0;
if delinq eq 1 then delinq0 = 1;
if delinq eq 2 then delinq1 = 1;
if delinq eq 3 then delinq2 = 1;
if delinq eq _missing then delinq0 = _missing;
if delinq eq _missing then delinq1 = _missing;
if delinq eq _missing then delinq2 = _missing;

grade_a = 0;
grade_b = 0;
grade_c = 0;
grade_d = 0;
if grades eq 1 then grade_a = 1;
if grades eq 2 then grade_b = 1;
if grades eq 3 then grade_c = 1;
if grades eq 4 then grade_d = 1;
if grades eq _missing then grade_a = _missing;
if grades eq _missing then grade_b = _missing;
if grades eq _missing then grade_c = _missing;
if grades eq _missing then grade_d = _missing;

encour0 = 0;
encour1 = 0;
encour2 = 0;
if encourag eq 0 then encour0 = 1;
if encourag eq 1 then encour1 = 1;
if encourag eq 2 then encour2 = 1;
if encourag eq _missing then encour0 = _missing;
if encourag eq _missing then encour1 = _missing;
if encourag eq _missing then encour2 = _missing;

rel0     = 0;
rel1_2   = 0;
rel3_5   = 0;
rel6_24  = 0;
rel25_52 = 0;
rel_gt53 = 0;
if religsrv eq 1 then rel0     = 1;
if religsrv eq 2 then rel1_2   = 1;
if religsrv eq 3 then rel3_5   = 1;
if religsrv eq 4 then rel6_24  = 1;
if religsrv eq 5 then rel25_52 = 1;
if religsrv eq 6 then rel_ge53 = 1;
if religsrv eq _missing then rel0     = _missing;
if religsrv eq _missing then rel1_2   = _missing;
if religsrv eq _missing then rel3_5   = _missing;
if religsrv eq _missing then rel6_24  = _missing;
if religsrv eq _missing then rel25_52 = _missing;
if religsrv eq _missing then rel_ge53 = _missing;
```

Exhibit D.1 SUDAAN® and Mplus® Syntax for Complete Case Ordered Multinomial Logistic Regression Analysis (continued)

```
roleimpr = 0;
noimpr = 0;
if mdeimpy eq 1 then roleimpr = 1;
if mdeimpy eq 2 then noimpr = 1;
if mdeimpy eq _missing then roleimpr = _missing;
if mdeimpy eq _missing then noimpr =   _missing;

mh_1    = 0;
mh_2    = 0;
mh_3_6  = 0;
mh_7_24 = 0;
mh_ge25 = 0;
mh_0    = 0;
if smhvst eq 1 then mh_1    = 1;
if smhvst eq 2 then mh_2    = 1;
if smhvst eq 3 then mh_3_6  = 1;
if smhvst eq 4 then mh_7_24 = 1;
if smhvst eq 5 then mh_ge25 = 1;
if smhvst eq 6 then mh_0    = 1;
if smhvst eq _missing then mh_1    = _missing;
if smhvst eq _missing then mh_2    = _missing;
if smhvst eq _missing then mh_3_6  = _missing;
if smhvst eq _missing then mh_7_24 = _missing;
if smhvst eq _missing then mh_ge25 = _missing;
if smhvst eq _missing then mh_0    = _missing;

yesmeds = 0;
nomeds  = 0;
if meds eq 1 then yesmeds = 1;
if meds eq 2 then nomeds  = 1;
if meds eq _missing then yesmeds = _missing;
if meds eq _missing then nomeds  = _missing;

! reverse code the outcome to ensure identical results to SUDAAN
yotmthlp = -1*yotmthlp + 6;
! end variable recoding

VARIABLE:
NAMES = vestr verep year questid yotmthlp age irsex race4 income5 irinsur4
rururb00 delinq grades encourag religsrv mdeimpy smhvst
meds analwt year2;

MISSING=.;

WEIGHT = analwt;
STRATIFICATION = vestr;
CLUSTER = verep;
SUBPOPULATION = yotmthlp NE _missing;

IDVAR = questid;
```

Exhibit D.1 SUDAAN® and Mplus® Syntax for Complete Case Ordered Multinomial Logistic Regression Analysis (continued)

```
USEVARIABLES ARE yotmthlp
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;

CATEGORICAL ARE yotmthlp;

ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR IS MLR;

MODEL:
YOTMTHLP ON
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;

OUTPUT: stdyx tech1;
```

Table D.1 Results of Complete Case Ordered Multinomial Logistic Regression Analyses

	SUDAAN® $n = 2,820$			Mplus® $n = 2,820$		
Parameter	**Beta**	**SE**	***P*-value**	**Beta**	**SE**	***P*-value**
Year: 2006 (vs. 2010)	-0.085	0.142	0.551	-0.085	0.142	0.551
Year: 2007 (vs. 2010)	0.047	0.141	0.737	0.047	0.141	0.737
Year: 2008 (vs. 2010)	0.149	0.142	0.294	0.149	0.142	0.294
Year: 2009 (vs. 2010)	-0.130	0.145	0.368	-0.130	0.145	0.368
Age: 12 (vs. 17)	-0.484	0.258	0.061	-0.484	0.258	0.061
Age: 13 (vs. 17)	0.011	0.184	0.954	0.011	0.184	0.954
Age: 14 (vs. 17)	0.152	0.162	0.348	0.152	0.162	0.348
Age: 15 (vs. 17)	-0.052	0.135	0.699	-0.052	0.135	0.699
Age: 16 (vs. 17)	0.033	0.133	0.803	0.033	0.133	0.803
Male (vs. Female)	-0.013	0.116	0.913	-0.013	0.116	0.913
Race: White (vs. Hispanic/Latino)	0.394	0.148	0.008	0.394	0.148	0.008
Race: Black/African American (vs. Hispanic/Latino)	0.067	0.197	0.734	0.067	0.197	0.734
Race: Other (vs. Hispanic/Latino)	0.378	0.233	0.105	0.378	0.233	0.105
Income: < $20k (vs. > $100k)	0.324	0.165	0.050	0.324	0.165	0.050
Income: $20-$50k (vs. > $100k)	0.148	0.132	0.262	0.148	0.132	0.262
Income: $50-$75k (vs. > $100k)	0.216	0.153	0.158	0.216	0.153	0.158
Income: $75-$100k (vs. > $100k)	0.379	0.157	0.016	0.379	0.157	0.016
Insured (vs. Uninsured)	0.311	0.191	0.103	0.311	0.191	0.103
Rural (vs. Urban)	0.031	0.101	0.759	0.031	0.101	0.759

Table D.1 Results of Complete Case Ordered Multinomial Logistic Regression Analyses (continued)

Parameter	SUDAAN® *n* = 2,820			Mplus® *n* = 2,820		
	Beta	SE	*P*-value	Beta	SE	*P*-value
0 Delinquent Acts (vs. 2+ Acts)	-0.588	0.114	0.000	-0.588	0.114	0.000
1 Delinquent Act (vs. 2+ Acts)	-0.540	0.127	0.000	-0.540	0.127	0.000
Grades: A (vs. D)	-0.285	0.163	0.080	-0.285	0.163	0.080
Grades: B (vs. D)	-0.314	0.143	0.028	-0.314	0.143	0.028
Grades: C (vs. D)	-0.060	0.154	0.696	-0.060	0.154	0.696
Parent Encouragement: None (vs. 2)	-0.147	0.165	0.000	-0.147	0.165	0.000
Parent Encouragement: 1 (vs. 2)	-0.626	0.118	0.001	-0.626	0.118	0.001
Religious Services: Never (vs. > Weekly)	0.480	0.152	0.002	0.480	0.152	0.002
Religious Services: 1-2/Year (vs. > Weekly)	0.331	0.178	0.062	0.331	0.178	0.062
Religious Services: 3-5/Year (vs. > Weekly)	0.290	0.185	0.116	0.290	0.185	0.116
Religious Services: 6-24/Year (vs. > Weekly)	0.337	0.160	0.036	0.337	0.160	0.036
Religious Services: Weekly (vs. > Weekly)	0.124	0.161	0.442	0.124	0.161	0.442
Role Impaired (vs. Unimpaired)	0.020	0.120	0.865	0.020	0.120	0.865
Past Year Mental Health Visit: 1 (vs. 0)	0.314	0.224	0.162	0.314	0.224	0.162
Past Year Mental Health Visit: 2 (vs. 0)	0.020	0.240	0.934	0.020	0.240	0.934
Past Year Mental Health Visit: 3-6 (vs. 0)	0.009	0.139	0.946	0.009	0.139	0.946
Past Year Mental Health Visit: 7-24 (vs. 0)	-0.165	0.130	0.205	-0.165	0.130	0.205
Past Year Mental Health Visit: 25+ (vs. 0)	-0.414	0.147	0.005	-0.414	0.147	0.005
Rx for Mood (vs. No Rx)	-0.382	0.105	0.000	-0.382	0.105	0.000

Rx = prescription; SE = standard error.

Note: Variables with missing data are shown in bold.

Note: The intercepts estimated in SUDAAN® and the thresholds estimated in Mplus® are not presented. Thresholds are an alternative parameterization of the intercepts and hence will not be identical to the intercepts.

D.4 FIML Analysis in Mplus

In the previous section, the equivalence of Mplus and SUDAAN results for complete case ordered multinomial logistic regression were established. In this section, how to address missing data using FIML in Mplus is discussed. Dependent variables with missing data already have a distributional assumption (e.g., they are ordered categorical as shown in Exhibit D.1). The *Mplus User's Guide* (Muthén & Muthén, 1998-2015) indicates "In all models, missingness is not allowed for the observed covariates because they are not part of the model. The outcomes are modeled conditional on the covariates and the covariates have no distributional assumption. Covariate missingness can be modeled if the covariates are explicitly brought into the model and given a distributional assumption." In the parlance of this appendix, a covariate is an independent variable.

Because missing data theory does not apply to exogenous (independent) variables, they must be brought into the distributional model (and not simply have the model conditioned on them). The way this is accomplished is to estimate the variances and covariances of the independent variables. All the independent variables need to be brought into the model. If any independent variables are excluded, even if they have no missingness, Mplus will assume zero covariances between covariates brought into the distributional model and those not brought into the distributional model. This assumption will rarely hold in practice. In Mplus, variances are estimated by adding the variable names with no special characters or keywords inside the MODEL command. Covariances are estimated by default, though this can be made explicit using the WITH keyword as shown in the following example (see Exhibit D.2 also):

```
! specify variances for all (binary) independent variable
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;

! specify covariances for all (binary) independent variable
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds
WITH
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;
```

As noted previously, the independent variables in this example are all binary. This is one advantage to explicitly defining each dummy coded variable rather than having an analog to the SAS and SUDAAN class statements. Specifically, the variance of a binary variable is more straightforward to work with than the variance of an ordered categorical variable (although ordered effect could be achieved by selecting alternative coding schemes to dummy coding). *Users should not put independent categorical variables in the CATEGORICAL subcommand of the VARIABLE command because Mplus will attempt to treat them as dependent variables and will terminate with an error when dependent variables have no predictors. This result makes explicit the fact that making distributional assumptions about a variable alone does not make it a dependent variable in the sense it is not conditioned on another variable (although it is dependent in the sense that a distribution is being specified for it).*

Independent variables brought into the model by requesting that their variances be estimated are assumed to be multivariate normal in addition to the assumed distribution for the dependent variable. In case of ordinal logistic regression, the distribution for the dependent variable is a multinomial distribution with an assumed underlying normal distribution and a link function (such as logit, the default in Mplus or probit). Note that binary logistic regression is a special case of ordinal logistic regression. As has been shown in the multiple imputation literature (Bernaards, Belin, & Schafer, 2007), assuming that binary variables are multivariate normal will still yield good results.

In this example, estimating the (co)variances of the independent variables required Monte Carlo integration. This is achieved by modifying the ANALYSIS command shown in Exhibit D.1 to look like this:

```
ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR IS MLR;
INTEGRATION IS MONTECARLO;
```

When a model cannot be estimated using standard integration methods, Mplus will terminate with an error and indicate that Monte Carlo integration is required. The FIML model was fit to the data by bringing all independent variables into the model as shown above.

Exhibit D.2 Mplus® Syntax for Full Information Maximum Likelihood Multinomial Logistic Regression Analysis

```
TITLE: logistic regression - dist on all vars - subpopn.inp
DATA: FILE = "S:/RTI Shares/NSDUH/FIMLguidance/Mplus/Data/mhrtr0610.dat";

! Use the define statement to create dummy variables (Mplus does not have a
! SAS/SUDAAN analog of a class statement); alternatively, this can be done
in
! another program such as SAS prior to exporting the data for Mplus
DEFINE:
y2006 = 0;
y2007 = 0;
y2008 = 0;
y2009 = 0;
y2010 = 0;
if year eq 2006 then y2006 = 1;
if year eq 2007 then y2007 = 1;
if year eq 2008 then y2008 = 1;
if year eq 2009 then y2009 = 1;
if year eq 2010 then y2010 = 1;

hlp_not  = 0;
hlp_lit  = 0;
hlp_some = 0;
hlp_alot = 0;
hlp_extr = 0;
if yotmthlp eq 1 then hlp_not  = 1;
if yotmthlp eq 2 then hlp_lit  = 1;
if yotmthlp eq 3 then hlp_some = 1;
if yotmthlp eq 4 then hlp_alot = 1;
if yotmthlp eq 5 then hlp_extr = 1;

age12 = 0;
age13 = 0;
age14 = 0;
age15 = 0;
age16 = 0;
age17 = 0;
if age eq 12 then age12 = 1;
if age eq 13 then age13 = 1;
if age eq 14 then age14 = 1;
if age eq 15 then age15 = 1;
if age eq 16 then age16 = 1;
if age eq 17 then age17 = 1;
```

Exhibit D.2 Mplus® Syntax for Full Information Maximum Likelihood Multinomial Logistic Regression Analysis (continued)

```
male   = 0;
female = 0;
if irsex eq 1 then male   = 1;
if irsex eq 2 then female = 1;

white = 0;
black = 0;
other = 0;
hisp  = 0;
if race4 eq 1 then white = 1;
if race4 eq 2 then black = 1;
if race4 eq 3 then other = 1;
if race4 eq 4 then hisp  = 1;

inlt20k  = 0;
in2050k  = 0;
in5075k  = 0;
in75100k = 0;
ingt100k = 0;
if income5 eq 1 then inlt20k  = 1;
if income5 eq 2 then in2050k  = 1;
if income5 eq 3 then in5075k  = 1;
if income5 eq 4 then in75100k = 1;
if income5 eq 5 then ingt100k = 1;

insure = 0;
uninsure = 0;
if irinsur4 eq 1 then insure = 1;
if irinsur4 eq 2 then uninsure = 1;

rural = 0;
urban = 0;
if rururb00 eq 1 then rural = 1;
if rururb00 eq 2 then urban = 1;

delinq0 = 0;
delinq1 = 0;
delinq2 = 0;
if delinq eq 1 then delinq0 = 1;
if delinq eq 2 then delinq1 = 1;
if delinq eq 3 then delinq2 = 1;
if delinq eq _missing then delinq0 = _missing;
if delinq eq _missing then delinq1 = _missing;
if delinq eq _missing then delinq2 = _missing;
```

Exhibit D.2 Mplus® Syntax for Full Information Maximum Likelihood Multinomial Logistic Regression Analysis (continued)

```
grade_a = 0;
grade_b = 0;
grade_c = 0;
grade_d = 0;
if grades eq 1 then grade_a = 1;
if grades eq 2 then grade_b = 1;
if grades eq 3 then grade_c = 1;
if grades eq 4 then grade_d = 1;
if grades eq _missing then grade_a = _missing;
if grades eq _missing then grade_b = _missing;
if grades eq _missing then grade_c = _missing;
if grades eq _missing then grade_d = _missing;

encour0 = 0;
encour1 = 0;
encour2 = 0;
if encourag eq 0 then encour0 = 1;
if encourag eq 1 then encour1 = 1;
if encourag eq 2 then encour2 = 1;
if encourag eq _missing then encour0 = _missing;
if encourag eq _missing then encour1 = _missing;
if encourag eq _missing then encour2 = _missing;

rel0     = 0;
rel1_2   = 0;
rel3_5   = 0;
rel6_24  = 0;
rel25_52 = 0;
rel_gt53 = 0;
if religsrv eq 1 then rel0     = 1;
if religsrv eq 2 then rel1_2   = 1;
if religsrv eq 3 then rel3_5   = 1;
if religsrv eq 4 then rel6_24  = 1;
if religsrv eq 5 then rel25_52 = 1;
if religsrv eq 6 then rel_ge53 = 1;
if religsrv eq _missing then rel0     = _missing;
if religsrv eq _missing then rel1_2   = _missing;
if religsrv eq _missing then rel3_5   = _missing;
if religsrv eq _missing then rel6_24  = _missing;
if religsrv eq _missing then rel25_52 = _missing;
if religsrv eq _missing then rel_ge53 = _missing;

roleimpr = 0;
noimpr = 0;
if mdeimpy eq 1 then roleimpr = 1;
if mdeimpy eq 2 then noimpr = 1;
if mdeimpy eq _missing then roleimpr = _missing;
if mdeimpy eq _missing then noimpr =    _missing;
```

Exhibit D.2 Mplus® Syntax for Full Information Maximum Likelihood Multinomial Logistic Regression Analysis (continued)

```
mh_1    = 0;
mh_2    = 0;
mh_3_6  = 0;
mh_7_24 = 0;
mh_ge25 = 0;
mh_0    = 0;
if smhvst eq 1 then mh_1    = 1;
if smhvst eq 2 then mh_2    = 1;
if smhvst eq 3 then mh_3_6  = 1;
if smhvst eq 4 then mh_7_24 = 1;
if smhvst eq 5 then mh_ge25 = 1;
if smhvst eq 6 then mh_0    = 1;
if smhvst eq _missing then mh_1    = _missing;
if smhvst eq _missing then mh_2    = _missing;
if smhvst eq _missing then mh_3_6  = _missing;
if smhvst eq _missing then mh_7_24 = _missing;
if smhvst eq _missing then mh_ge25 = _missing;
if smhvst eq _missing then mh_0    = _missing;

yesmeds = 0;
nomeds  = 0;
if meds eq 1 then yesmeds = 1;
if meds eq 2 then nomeds  = 1;
if meds eq _missing then yesmeds = _missing;
if meds eq _missing then nomeds  = _missing;

! reverse code the outcome to ensure identical results to SUDAAN
yotmthlp = -1*yotmthlp + 6;
! end variable recoding

VARIABLE:
NAMES = vestr verep year questid yotmthlp age irsex race4 income5 irinsur4
rururb00 delinq grades encourag religsrv mdeimpy smhvst meds analwt year2;

MISSING=.;

WEIGHT = analwt;
STRATIFICATION = vestr;
CLUSTER = verep;
SUBPOPULATION = yotmthlp NE _missing;

IDVAR = questid;

USEVARIABLES ARE yotmthlp
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;
```

Exhibit D.2 Mplus® Syntax for Full Information Maximum Likelihood Multinomial Logistic Regression Analysis (continued)

```
CATEGORICAL ARE yotmthlp;

ANALYSIS: TYPE IS COMPLEX;
ESTIMATOR IS MLR;
INTEGRATION IS MONTECARLO;
PROCESSORS ARE 8;

MODEL:
YOTMTHLP ON
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;


! specify variances for all (binary) independent variable
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;

! specify covariances for all (binary) independent variable
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds
WITH
y2006 y2007 y2008 y2009 age12 age13 age14 age15 age16
male white black other inlt20k in2050k in5075k in75100k insure
rural delinq0 delinq1 grade_a grade_b grade_c encour0 encour1
rel0 rel1_2 rel3_5 rel6_24 rel25_52 roleimpr
mh_1 mh_2 mh_3_6 mh_7_24 mh_ge25 yesmeds;

OUTPUT: stdyx tech1;
```

D.5 Results

The results of the FIML model are compared with the complete case model in Table D.2. To better illustrate the consequences of using a complete case model (i.e., assuming data were missing completely at random and ignoring missingness), the relative bias is computed to compare the two models. This is shown in Table D.3, where the relative bias of beta, the relative bias of the standard error, and an indicator of change in inference (alpha = 0.05) are shown for comparing complete case analysis with the FIML model.

Relative bias and change in inference should be considered together. For example, relative to 17-year-olds, 12-year-olds were less likely to report that counseling helped them in both the complete case model (beta = -0.484) and the FIML model (beta = -0.455). The complete case analysis represents a 6.4 percent positive bias in beta and a 15.2 percent positive bias in its standard error. However, the *p*-value was not significant under complete case analysis. Ignoring missingness would result in a type II error in statistical inference for this parameter. This example can be contrasted with the effect of being male, which has a negative coefficient bias over 1,000 percent, but does not impact the inference that there are no gender differences in reporting how much counseling helped.

Table D.4 presents the odds ratios and relative bias in odds ratios. Because of the nature of the transformation, the magnitudes of relative bias are often much less dramatic for the odds ratio than the beta coefficients.

Table D.2 FIML and Complete Case Model Results

	Complete Case Estimated Parameters = 42 *n* = 2,820 Run Time = 0:09			FIML Estimated Parameters = 821 *n* = 3,271 Run Time = 34:50		
Parameter	**Beta**	**SE**	***P*-value**	**Beta**	**SE**	***P*-value**
Year: 2006 (vs. 2010)	-0.085	0.142	0.551	-0.092	0.133	0.491
Year: 2007 (vs. 2010)	0.047	0.141	0.737	0.042	0.130	0.748
Year: 2008 (vs. 2010)	0.149	0.142	0.294	0.132	0.128	0.303
Year: 2009 (vs. 2010)	-0.130	0.145	0.368	-0.138	0.135	0.307
Age: 12 (vs. 17)	-0.484	0.258	0.061	-0.455	0.224	0.042
Age: 13 (vs. 17)	0.011	0.184	0.954	0.004	0.157	0.981
Age: 14 (vs. 17)	0.152	0.162	0.348	0.167	0.147	0.257
Age: 15 (vs. 17)	-0.052	0.135	0.699	0.001	0.127	0.996
Age: 16 (vs. 17)	0.033	0.133	0.803	0.040	0.127	0.753
Male (vs. Female)	-0.013	0.116	0.913	-0.003	0.109	0.977
Race: White (vs. Hispanic/Latino)	0.394	0.148	0.008	0.334	0.131	0.011
Race: Black/African American (vs. Hispanic/Latino)	0.067	0.197	0.734	0.028	0.177	0.875
Race: Other (vs. Hispanic/Latino)	0.378	0.233	0.105	0.297	0.209	0.155
Income: < $20k (vs. > $100k)	0.324	0.165	0.050	0.253	0.154	0.101
Income: $20-$50k (vs. > $100k)	0.148	0.132	0.262	0.074	0.124	0.550
Income: $50-$75k (vs. > $100k)	0.216	0.153	0.158	0.156	0.143	0.272
Income: $75-$100k (vs. > $100k)	0.379	0.157	0.016	0.300	0.147	0.041
Insured (vs. Uninsured)	0.311	0.191	0.103	0.277	0.185	0.134
Rural (vs. Urban)	0.031	0.101	0.759	0.019	0.090	0.829
0 Delinquent Acts (vs. 2+ Acts)	-0.588	0.114	0.000	-0.569	0.106	0.000
1 Delinquent Act (vs. 2+ Acts)	-0.540	0.127	0.000	-0.509	0.120	0.000
Grades: A (vs. D)	-0.285	0.163	0.080	-0.329	0.159	0.038
Grades: B (vs. D)	-0.314	0.143	0.028	-0.339	0.140	0.015
Grades: C (vs. D)	-0.060	0.154	0.696	-0.131	0.149	0.377
Parent Encouragement: None (vs. 2)	-0.147	0.165	0.000	0.586	0.111	0.000
Parent Encouragement: 1 (vs. 2)	-0.626	0.118	0.001	0.426	0.131	0.001
Religious Services: Never (vs. > Weekly)	0.480	0.152	0.002	0.502	0.142	0.000
Religious Services: 1-2/Year (vs. > Weekly)	0.331	0.178	0.062	0.352	0.168	0.036
Religious Services: 3-5/Year (vs. > Weekly)	0.290	0.185	0.116	0.323	0.174	0.064
Religious Services: 6-24/Year (vs. > Weekly)	0.337	0.160	0.036	0.330	0.153	0.030
Religious Services: Weekly (vs. > Weekly)	0.124	0.161	0.442	0.206	0.154	0.180
Role Impaired (vs. Unimpaired)	0.020	0.120	0.865	0.046	0.112	0.683

Table D.2 FIML and Complete Case Model Results (continued)

Parameter	Complete Case Estimated Parameters = 42 n = 2,820 Run Time = 0:09			FIML Estimated Parameters = 821 n = 3,271 Run Time = 34:50		
	Beta	SE	*P*-value	Beta	SE	*P*-value
Past Year Mental Health Visit: 1 (vs. 0)	0.314	0.224	0.162	0.207	0.220	0.347
Past Year Mental Health Visit: 2 (vs. 0)	0.020	0.240	0.934	-0.032	0.231	0.889
Past Year Mental Health Visit: 3-6 (vs. 0)	0.009	0.139	0.946	-0.039	0.132	0.765
Past Year Mental Health Visit: 7-24 (vs. 0)	-0.165	0.130	0.205	-0.194	0.126	0.123
Past Year Mental Health Visit: 25+ (vs. 0)	-0.414	0.147	0.005	-0.468	0.142	0.001
Rx for Mood (vs. No Rx)	-0.382	0.105	0.000	-0.366	0.099	0.000

FIML = full information maximum likelihood; Rx = prescription; SE = standard error.
Note: Variables with missing data are shown in bold.
Note: Run times were computed by Mplus®. Mplus version 7.2 64-bit was run on a Windows 7 Professional laptop with a quad-core 2.7 GHz processor and 8GB RAM.

Table D.3 Relative Bias of Complete Case Results Compared with FIML

Parameter	Beta	SE	Error Type
Year: 2006 (vs. 2010)	-7.6%	6.8%	
Year: 2007 (vs. 2010)	11.9%	8.5%	
Year: 2008 (vs. 2010)	12.9%	10.9%	
Year: 2009 (vs. 2010)	-5.8%	7.4%	
Age: 12 (vs. 17)	6.4%	15.2%	II
Age: 13 (vs. 17)	175.0%	17.2%	
Age: 14 (vs. 17)	-9.0%	10.2%	
Age: 15 (vs. 17)	-5,300.0%	6.3%	
Age: 16 (vs. 17)	-17.5%	4.7%	
Male (vs. Female)	333.3%	6.4%	
Race: White (vs. Hispanic/Latino)	18.0%	13.0%	
Race: Black/African American (vs. Hispanic/Latino)	139.3%	11.3%	
Race: Other (vs. Hispanic/Latino)	27.3%	11.5%	
Income: < \$20k (vs. > \$100k)	28.1%	7.1%	I
Income: \$20-\$50k (vs. > \$100k)	100.0%	6.5%	
Income: \$50-\$75k (vs. > \$100k)	38.5%	7.0%	
Income: \$75-\$100k (vs. > \$100k)	26.3%	6.8%	
Insured (vs. Uninsured)	12.3%	3.2%	
Rural (vs. Urban)	63.2%	12.2%	
0 Delinquent Acts (vs. 2+ Acts)	3.3%	7.5%	
1 Delinquent Act (vs. 2+ Acts)	6.1%	5.8%	
Grades: A (vs. D)	-13.4%	2.5%	II
Grades: B (vs. D)	-7.4%	2.1%	
Grades: C (vs. D)	-54.2%	3.4%	
Parent Encouragement: None (vs. 2)	-125.1%	48.6%	
Parent Encouragement: 1 (vs. 2)	-246.9%	-9.9%	
Religious Services: Never (vs. > Weekly)	-4.4%	7.0%	
Religious Services: 1-2/Year (vs. > Weekly)	-6.0%	6.0%	II
Religious Services: 3-5/Year (vs. > Weekly)	-10.2%	6.3%	
Religious Services: 6-24/Year (vs. > Weekly)	2.1%	4.6%	
Religious Services: Weekly (vs. > Weekly)	-39.8%	4.5%	
Role Impaired (vs. Unimpaired)	-56.5%	7.1%	

Table D.3 Relative Bias of Complete Case Results Compared with FIML (continued)

Parameter	Beta	SE	Error Type
Past Year Mental Health Visit: 1 (vs. 0)	51.7%	1.8%	
Past Year Mental Health Visit: 2 (vs. 0)	-162.5%	3.9%	
Past Year Mental Health Visit: 3-6 (vs. 0)	-123.1%	5.3%	
Past Year Mental Health Visit: 7-24 (vs. 0)	-14.9%	3.2%	
Past Year Mental Health Visit: 25+ (vs. 0)	-11.5%	3.5%	
Rx for Mood (vs. No Rx)	4.4%	6.1%	

FIML = full information maximum likelihood; Rx = prescription; SE = standard error.

Note: Variables with missing data are shown in bold.

Note: Relative bias is computed as 100*(CC – FIML)/FIML, where CC indicates complete cases analysis results. The error type column indicates whether a type I or type II error was made in the complete case analysis relative to FIML analysis (alpha = 0.05). Empty cells indicate no change in inference for a given parameter.

Table D.4 Odds Ratio and Relative Bias in Odds Ratio Results for Complete Case and FIML Models

	Odds Ratio		
Parameter	**Complete Case**	**FIML**	**Relative Bias of Odds Ratio**
Year: 2006 (vs. 2010)	0.919	0.912	0.8%
Year: 2007 (vs. 2010)	1.048	1.043	0.5%
Year: 2008 (vs. 2010)	1.161	1.141	1.8%
Year: 2009 (vs. 2010)	0.878	0.871	0.8%
Age: 12 (vs. 17)	0.616	0.634	-2.8%
Age: 13 (vs. 17)	1.011	1.004	0.7%
Age: 14 (vs. 17)	1.165	1.181	-1.4%
Age: 15 (vs. 17)	0.949	1.001	-5.2%
Age: 16 (vs. 17)	1.034	1.041	-0.7%
Male (vs. Female)	0.987	0.997	-1.0%
Race: White (vs. Hispanic/Latino)	1.483	1.396	6.2%
Race: Black/African American (vs. Hispanic/Latino)	1.069	1.028	4.0%
Race: Other (vs. Hispanic/Latino)	1.460	1.346	8.5%
Income: < $20k (vs. > $100k)	1.383	1.288	7.4%
Income: $20-$50k (vs. > $100k)	1.160	1.077	7.7%
Income: $50-$75k (vs. > $100k)	1.241	1.169	6.2%
Income: $75-$100k (vs. > $100k)	1.461	1.350	8.2%
Insured (vs. Uninsured)	1.365	1.319	3.5%
Rural (vs. Urban)	1.032	1.020	1.2%
0 Delinquent Acts (vs. 2+ Acts)	0.556	0.566	-1.8%
1 Delinquent Act (vs. 2+ Acts)	0.583	0.601	-3.0%
Grades: A (vs. D)	0.752	0.719	4.6%
Grades: B (vs. D)	0.731	0.712	2.7%
Grades: C (vs. D)	0.942	0.877	7.4%
Parent Encouragement: None (vs. 2)	1.870	1.798	4.0%
Parent Encouragement: 1 (vs. 2)	1.615	1.531	5.5%
Religious Services: Never (vs. > Weekly)	1.617	1.651	-2.1%
Religious Services: 1-2/Year (vs. > Weekly)	1.393	1.422	-2.0%
Religious Services: 3-5/Year (vs. > Weekly)	1.336	1.381	-3.3%
Religious Services: 6-24/Year (vs. > Weekly)	1.400	1.391	0.6%
Religious Services: Weekly (vs. > Weekly)	1.131	1.229	-8.0%
Role Impaired (vs. Unimpaired)	1.021	1.047	-2.5%

Table D.4 Odds Ratio and Relative Bias in Odds Ratio Results for Complete Case and FIML Models (continued)

Parameter	Odds Ratio: Complete Case	Odds Ratio: FIML	Relative Bias of Odds Ratio
Past Year Mental Health Visit: 1 (vs. 0)	1.369	1.230	11.3%
Past Year Mental Health Visit: 2 (vs. 0)	1.020	0.968	5.4%
Past Year Mental Health Visit: 3-6 (vs. 0)	1.009	0.961	5.0%
Past Year Mental Health Visit: 7-24 (vs. 0)	0.848	0.824	2.9%
Past Year Mental Health Visit: 25+ (vs. 0)	0.661	0.626	5.6%
Rx for Mood (vs. No Rx)	0.683	0.693	-1.4%

FIML = full information maximum likelihood; Rx = prescription; SE = standard error.

Note: Variables with missing data are shown in bold.

Note: Relative bias is computed as 100*(CC – FIML)/FIML, where CC indicates complete cases analysis results.

Appendix E: How to Implement Full Information Maximum Likelihood Logistic Regression in Latent GOLD®

E.1 Introduction

This appendix describes the full information maximum likelihood (FIML) approach using complex survey data in the Latent GOLD® software package. To conduct this analysis using the Latent GOLD software, the basic, advanced, and syntax packages must be installed on the machine.

Included is a demonstration for complete case analysis, showing that the SUDAAN®[51] and Latent GOLD software programs give similar results. Due to Latent GOLD's variance calculation method, it is recommended that Latent GOLD be used only for models with none to very few single-unit strata. Latent GOLD treats single-unit strata as certainty strata by default, which may underestimate the variance. Although this is not discussed below, it can be addressed by altering the dataset to eliminate single-unit strata by collapsing them together.[52]

For this demonstration, the ordered categorical (ordinal) outcome variable YOTMTHLP in the National Survey on Drug Use and Health (NSDUH) is modeled using ordered multinomial logistic regression. This variable is on a 5-point Likert-type scale indicating whether counseling helped the respondent (i.e., not at all, a little, some, a lot, or extremely). The independent variables include survey year, age, gender, race/ethnicity, family income, health insurance status, rural/urban domicile, number of past year delinquent behaviors, grades in school, parental encouragement, number of religious services attended in the past year, severe role impairment, number of mental health visits in the past year, and the status of a prescription for mood medication in the past year.

E.2 FIML in Latent GOLD

Latent GOLD has two modes of user interaction: (1) a graphical user interface (GUI), and (2) a syntax file. To address missing data using the FIML approach on the independent variables, a syntax file must be used. Latent GOLD can generate a base syntax file to which modifications can be made to allow for FIML estimation. Following are characteristics of the Latent GOLD syntax language:

- The comment character can be "//" or "/*" and "*/".
- Commands end with a semicolon (";").
- It reads in SAV data files.
- The Latent GOLD User's Guide, Upgrade Manual, and Technical Guide can be downloaded from http://www.statisticalinnovations.com/user-guides/.

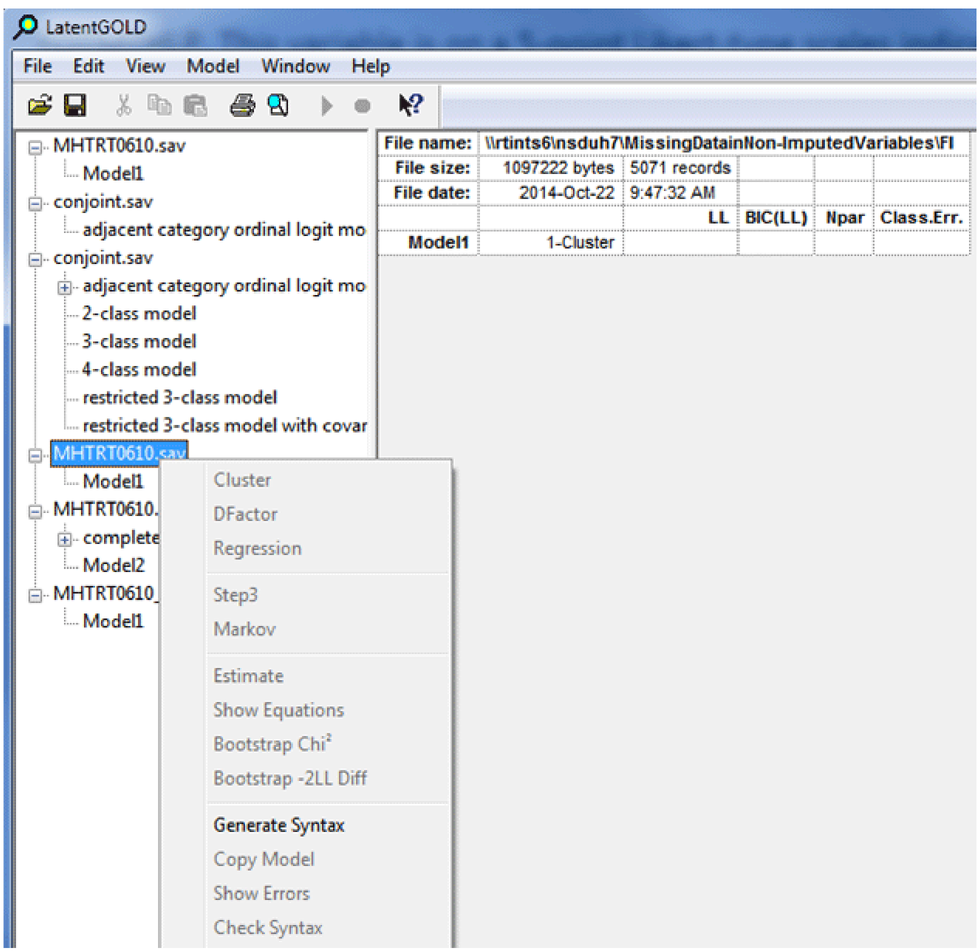
[51] SUDAAN® is a registered trademark of Research Triangle Institute.

[52] Although not tested for this analysis, as of version 5.1, Latent GOLD now handles single-unit strata with certainty, scaled, or centered options. To use these options, add the keyword (certainty, scaled, or centered) after the stratumid variable.

E.3 Creating a Base Syntax File

To create a base syntax file, first open the data file in Latent GOLD. Latent GOLD reads in SAV data files. The SAS®[53] dataset may need to be exported into SAV format using PROC EXPORT in the SAS system. Also, if any variables require modification prior to model estimation in Latent GOLD, these modifications should be made in SAS. For example, the YOTMTHLP variable was recoded in SAS and then exported to a SAV file for use in Latent GOLD. This was done to ensure the model output from Latent GOLD matched that of the SUDAAN output. The data file read into Latent GOLD should only contain the observations that will be included in the model. For example, the YOTMTHLP variable is only asked of youths, so any adults on the original SAS dataset were removed from the SAV dataset that was used in Latent GOLD to model YOTMTHLP.

Once opened, the name of the data file with Model1 in the left-hand panel will be visible, as shown in the screenshot below. Right click on the dataset and select "Generate Syntax."[54] Latent GOLD syntax files are composed of three parts: options, variables, and equations.



[53] SAS® software is a registered trademark of SAS Institute Inc.
[54] The following warning message may pop up: No Dependent Variables defined! in line 17: Error in equations: "end-of-file". Ignore this warning by clicking "OK" to close the window.

The basic syntax file generated by Latent GOLD is shown below.

```
1     options
2        maxthreads=all;
3        algorithm
4           tolerance=1e-008 emtolerance=0.01 emiterations=250
            nriterations=50 ;
5        startvalues
6           seed=0 sets=16 tolerance=1e-005 iterations=50;
7        bayes
8           categorical=1 variances=1 latent=1 poisson=1;
9        montecarlo
10          seed=0 sets=0 replicates=500 tolerance=1e-008;
11       quadrature  nodes=10;
12       missing  excludeall;
13       output
14          parameters=effect  betaopts=wl standarderrors profile
            probmeans=posterior
15          bivariateresiduals estimatedvalues=regression;
16    variables
17    equations
```

E.4 Editing Syntax Files for FIML, OPTIONS Section

The OPTIONS section allows the user to specify the estimation parameters, describes how to handle missing data, and allows the user to request desired outputs. Unless the model fails to converge, the estimation parameters should not require modification. In the case of nonconvergence, increasing the number of iterations might help find the global maximum.

The "missing" statement on line 12 above has two options: "includeall" and "excludeall." Excludeall is the equivalent of a complete case analysis. When excludeall is requested, an observation with any independent or dependent variable (as specified in the VARIABLES section) with a missing value is removed from the analysis. When includeall is requested, all observations are used in the analysis. By default, Latent GOLD will use FIML on the independent variables and a mean imputation method on the dependent variables.

Because the built-in FIML estimation is to be used, the missing command will need to be edited to say "includeall." (*Quasi* latent variables will be used to "trick" Latent GOLD into performing FIML on the independent variables. See the VARIABLES section for details.)

Next, set up the output statement on lines 13 to 15. By default, Latent GOLD requests several outcomes in the output section including *parameters, betaopts, standarderrors, profile, probmeans, bivariateresiduals,* and *estimatedvalues*. Many of these are useful outputs for latent variable models.[55] For fitting a regression model, the following output statement is sufficient.

```
output
  parameters=last standarderrors estimatedvalues estimatedvalues=regression;
```

[55] *Profile* provides information on class sizes and class-specific response probabilities/means/rates, *probmeans* displays the average posterior probabilities/means for group-level and dynamic latent variables, and *bivariateresiduals* produces output to detect local independence in latent class analysis.

Latent GOLD can handle effect and reference cell coding for the parameter estimates; effect coding is the default. To request reference cell coding, specify "parameters = last" or "parameters = first," depending on whether Latent GOLD will use the first or the last category as the reference. Beta estimates for every model specified will output in the "Parameters" suboption below the model specification.

The "standarderrors" command will produce robust standard error estimates in the context of complex survey data.

The "estimatedvalues" and "estimatedvalues=regression" options output model likelihood estimates for every cell in the classification table in two different structures. The "estimatedvalues=regression" option provides model likelihood estimates with the dependent variable as part of the classification table (i.e., a row variable); the "estimatedvalues" option uses only the independent variables for the classification table (i.e., a column variable). In both outputs, the explanatory variables are displayed as row variables. The beta estimates for every model will be output in the "Parameters" suboption.

E.5 Editing Syntax Files for FIML, VARIABLES Section

The VARIABLES section is the space where the complex data structure ("stratumID," "psuid," and "samplingweight") and independent ("independent"), dependent ("dependent"), and quasi-latent ("latent") variables that will be used in the regression model will be defined.

Below is the VARIABLES section defined to model the YOTMTHLP variable.

```
variables
   stratumID vestr;                  // Complex Survey Design
   psuid verep;
   samplingweight analwt rescale;
   dependent yotmthlp_2 cumlogit;    // Modeling YOTMHTLP_2 with
                                     //   a cumulative logit function
     // Explanatory variables defined by variable name and data type
   independent year nominal, age nominal, irsex nominal, income5 nominal,
               irinsur4 nominal, rururb00 nominal,
               delinquency nominal, grades nominal, encouragement nominal,
               religServices nominal, mdeimpy nominal, smhvst nominal,
               meds nominal, race4 nominal ;
   latent
      q_grades nominal 4, q_delinquency nominal 3,
      q_encouragement nominal 3, q_religServices nominal 6,
      q_mdeimpy nominal 2, q_smhvst nominal 6, q_meds nominal 2 ;
```

To specify the complex data structure within NSDUH, the "stratumID" should be vestr, the "psuid" should be verep, and the "samplingweight" should be analwt with the rescale option. The rescale option adjusts the original weights by multiplying them by a constant so that they sum up to the sample size.

The dependent variable is the variable to be modeled and should have the form:

dependent *<variable name> <type of model>*

For this example, the dependent variable is YOTMTHLP. Note that the "_2" is due to a recoding in SAS to change the order of the variable that was used to match results from SUDAAN. After specifying the dependent variable, the type of model to be fit should then be specified. In the example, a cumulative logistic model is fit. Other models that can be fit are presented below with their keywords.

Linear Normal	continuous
Multinomial Logistic	nominal
Cumulative Logistic	cumlogit
Adjacent Category Ordinal Logistic	ordinal (*this is the default*)
Log-Linear Poisson	poisson
Binary Logistic for Counts	binomial
Ordinal Probit	probit
Ordinal Log-Log	loglog1 or loglog2
Sequential Logit	seqlogit1 or seqlogit2
Log-Linear Gamma	gamma
Logit Beta	beta
Linear Von Mises	vonmises

For independent variables, Latent GOLD must be informed of the variable type (either nominal or numeric) as well as the variable name. Multiple independent variables are specified on the same line and should be separated by a comma (","). The basic independent line should have the form:

independent *<variable name> <variable type>, <variable name>*
<variable type>, . . . , <variable name> <variable type>

The last line is for latent variables ("latent"). This section is required to "trick" Latent GOLD into performing FIML estimation on the independent variables with missingness. For each independent variable with missing values for which FIML is required, a "q_" version of this variable is specified on the "latent" line. Latent variables have three parts: a name, a type, and, if the type is nominal, the number of categories.

The latent line should have the following basic form:

latent q_*<variable name> <variable type> <# categories>,*
q_*<variable name> <variable type> <# categories>, . . . ,*
q_*<variable name> <variable type> <# categories>*

Note that if there are variables in the model with missing data that will not need FIML estimation, then Latent GOLD will by default fill in the missing values with a mean imputation. In Version 5.0, there is no way to turn this off. Therefore, it is recommended that the dataset be reduced to include only complete cases for those variables not requiring FIML.

E.6 Editing Syntax Files for FIML, EQUATIONS Section

Now that the OPTIONS and VARIABLES sections are defined, the equations that will be used to create the regression model can be defined. There are three types of equations: (1) latent variable models, (2) weight equations, and (3) dependent variable models.

In Latent GOLD, latent variables can be placed on either the left-hand or right-hand side of an equation. By placing a latent variable on the left-hand side of the equation, Latent GOLD will use FIML estimation on those observations with missing values. The latent variable model has the following basic form:

<*quasi latent variable*> <- (w1~wei) <*original independent variable*>

In the statement above, w1 is a weight equation; Latent GOLD knows to treat w1 as a weight equation because of the special parameter "~wei" or "~weight." This weight statement is required in the latent variable model to ensure a 1:1 mapping of known values from the original variable to the *quasi* latent variable. Below are examples of weight statements for common nominal variables.

Number of Levels	Weight Statement
2	w1 = { 1 0 0 1};
3	w1 = {1 0 0 0 1 0 0 0 1};
4	w1 = {1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1};

Together the weight equation and the latent variable equation end up having the following form:

<*quasi latent variable*> <- (w1~wei) <*original independent variable*>*;*
w1 = {1 0   0 1}; // assuming the original variable has 2 categories

Last, the equation for the dependent variable uses a mixture of the *quasi* latent variables and independent variables. If a *quasi* latent variable was created for an independent variable, then the *quasi* latent variable will be used in the model instead of the original independent variable. Remember the *quasi* latent variable has been estimated using FIML. The dependent variable model takes the following form with the understanding that *quasi* latent and independent variables can be listed in any order:

<*dependent variable*> <- 1 + <*independent variables*> + <*quasi latent variables*>

The final syntax file for the cumulative logistic estimation of YOTMHTLP using FIML on the independent variables with missing data is shown in Exhibit E.1.

Exhibit E.1 Final Syntax for Cumulative Logistic Regression Analysis

```
options
   maxthreads=all;
   algorithm
      tolerance=1e-008 emtolerance=0.01 emiterations=250 nriterations=50 ;
   startvalues
      seed=0 sets=16 tolerance=1e-005 iterations=50;
   bayes
      categorical=1 variances=1 latent=1 poisson=1;
   montecarlo
      seed=0 sets=0 replicates=500 tolerance=1e-008;
   quadrature  nodes=10;
   missing  includeall;
   output
      parameters=last standarderrors estimatedvalues
       estimatedvalues=regression;

variables
   stratumID vestr;
   psuid verep;
   samplingweight analwt rescale;
   dependent yotmthlp_2 cumlogit;
   independent year nominal, age nominal, irsex nominal, income5 nominal,
               irinsur4 nominal, rururb00 nominal,
               delinquency nominal, grades nominal, encouragement nominal,
               religServices nominal, mdeimpy nominal,
               smhvst nominal, meds nominal, race4 nominal ;
   latent q_grades nominal 4, q_delinquency nominal 3,
          q_encouragement nominal 3, q_religServices nominal 6,
          q_mdeimpy nominal 2, q_smhvst nominal 6, q_meds nominal 2;

equations

  q_grades <- (w1~wei) grades;
  q_delinquency <- (w2~wei) delinquency ;
  q_encouragement  <- (w2~wei) encouragement ;
  q_mdeimpy <- (w3~wei) medimpy;
  q_meds <- (w3~wei) meds;
  q_religServices  <- (w4~wei) religServices ;
  q_smhvst <- (w4~wei) smhvst;

   yotmthlp_2 <- 1 + year + age + irsex + race4 + income5 + irinsur4 +
            rururb00 + q_delinquency + q_grades + q_encouragement +
            religServices + q_mdeimpy + smhvst + q_meds;

   w1 = {1 0 0 0      0 1 0 0      0 0 1 0      0 0 0 1};
   w2 = {1 0 0        0 1 0        0 0 1};
   w3 = {1 0          0 1};
   w4 = {1 0 0 0 0 0  0 1 0 0 0 0  0 0 1 0 0 0  0 0 0 1 0 0  0 0 0 0 1 0
         0 0 0 0 0 1};
```

This page intentionally left blank

Appendix F: Results of the Simulation Experiment for Deletion Rates of 5 Percent, 12.5 Percent, and 20 Percent

This appendix contains the following tables:

- Tables F.1 to F.14: These tables are analogous to Tables 4.11 to 4.17 in Chapter 4 but for 5 percent and 12.5 percent deletion rates instead of 20 percent. For example, Table F.1 is analogous to Table 4.11 but for the 5 percent deletion rate; Table F.2 is analogous to Table 4.11 but for the 12.5 percent deletion rate; Table F.3 is analogous to Table 4.12 but for the 5 percent deletion rate; Table F.4 is analogous to Table 4.12 but for the 12.5 percent deletion rate; and so on. Briefly, these tables assess the performance of the missing item values in regression analyses (MIVRA) method with respect to the bias and variance of the estimates of the regression coefficients, and they generally show the same results as Tables 4.11 to 4.17 but to a lesser extent.
- Tables F.15 and F.16: These tables are analogous to Table 4.18 but for 5 percent and 12.5 percent deletion rates instead of 20 percent. Briefly, these tables assess whether the MIVRA method matches the completed sample with respect to the statistical significance of the regression coefficients. Generally, the match rate for the MIVRA method decreases as the deletion rate increases, as expected.
- Tables F.17 to F.21: These tables are analogous to Table 4.19 but for the other five analytic models. Table 4.19 displays statistics for the N19/ANLYR analytic model only. These five tables tend to show the same results as Table 4.19; that is, the match rate for each regression coefficient depends mainly on how close the completed-sample *p*-value of the regression coefficient is to 0.05.

Table F.1 Weighted Distribution of Absolute Empirical Relative Biases by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute Empirical Relative Bias LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	0.001	0.000	0.000	0.001	0.000
	1st Quartile	0.003	0.001	0.003	0.010	0.002
	Median	0.007	0.002	0.005	0.012	0.003
	Mean	0.012	0.008	0.006	0.031	0.005
	3rd Quartile	0.011	0.013	0.009	0.045	0.008
	Maximum	5.299	2.015	0.465	1.958	0.801
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	0.001	0.001	0.000	0.000	0.000
	1st Quartile	0.005	0.002	0.003	0.001	0.001
	Median	0.014	0.006	0.009	0.001	0.003
	Mean	0.024	0.013	0.013	0.003	0.006
	3rd Quartile	0.033	0.015	0.013	0.002	0.005
	Maximum	3.554	1.141	0.342	0.103	0.323
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	0.001	0.000	0.000	0.000	0.000
	1st Quartile	0.003	0.001	0.001	0.000	0.001
	Median	0.006	0.002	0.003	0.001	0.003
	Mean	0.009	0.005	0.009	0.003	0.009
	3rd Quartile	0.010	0.006	0.012	0.003	0.010
	Maximum	0.352	0.108	0.258	0.328	0.346
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.001	0.000	0.000	0.000	0.000
	1st Quartile	0.004	0.001	0.002	0.001	0.002
	Median	0.008	0.003	0.005	0.002	0.004
	Mean	0.014	0.005	0.009	0.004	0.010
	3rd Quartile	0.018	0.007	0.016	0.003	0.020
	Maximum	0.121	0.042	0.065	0.097	0.129
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	0.000	0.000	0.000	N/A	0.000
	1st Quartile	0.001	0.000	0.005	N/A	0.003
	Median	0.002	0.001	0.008	N/A	0.006
	Mean	0.008	0.004	0.011	N/A	0.009
	3rd Quartile	0.014	0.002	0.016	N/A	0.014
	Maximum	0.167	0.042	0.051	N/A	0.197
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	0.001	0.000	0.004	0.000	0.000
	1st Quartile	0.002	0.001	0.008	0.000	0.003
	Median	0.008	0.002	0.012	0.001	0.006
	Mean	0.014	0.005	0.016	0.005	0.013
	3rd Quartile	0.013	0.004	0.021	0.004	0.020
	Maximum	0.138	0.076	0.063	0.087	0.057

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $|relB|$ from equation 4.5, with the coefficients' $|c|$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., .032 is 3.2%).

Note: See Table 4.1 for study and model details.

Table F.2 Weighted Distribution of Absolute Empirical Relative Biases by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute Empirical Relative Bias				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	0.002	0.000	0.001	0.003	0.000
	1st Quartile	0.007	0.001	0.008	0.021	0.004
	Median	0.018	0.007	0.012	0.025	0.006
	Mean	0.032	0.024	0.014	0.065	0.012
	3rd Quartile	0.039	0.040	0.021	0.098	0.020
	Maximum	14.630	6.168	0.909	4.286	1.549
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	0.000	0.004	0.001	0.000	0.000
	1st Quartile	0.012	0.007	0.007	0.002	0.002
	Median	0.037	0.019	0.021	0.003	0.004
	Mean	0.068	0.040	0.032	0.007	0.015
	3rd Quartile	0.093	0.050	0.033	0.005	0.015
	Maximum	10.147	3.923	0.526	0.266	0.697
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	0.000	0.000	0.000	0.000	0.001
	1st Quartile	0.005	0.003	0.002	0.001	0.004
	Median	0.015	0.005	0.007	0.002	0.009
	Mean	0.022	0.012	0.022	0.008	0.024
	3rd Quartile	0.023	0.015	0.020	0.006	0.024
	Maximum	1.516	0.615	0.741	0.788	0.746
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.002	0.000	0.001	0.000	0.000
	1st Quartile	0.010	0.006	0.003	0.001	0.004
	Median	0.022	0.010	0.012	0.005	0.011
	Mean	0.042	0.018	0.022	0.012	0.024
	3rd Quartile	0.048	0.021	0.042	0.012	0.048
	Maximum	0.367	0.124	0.198	0.293	0.243
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	0.000	0.000	0.000	N/A	0.001
	1st Quartile	0.003	0.001	0.013	N/A	0.010
	Median	0.009	0.001	0.019	N/A	0.012
	Mean	0.023	0.012	0.029	N/A	0.025
	3rd Quartile	0.043	0.007	0.040	N/A	0.039
	Maximum	0.495	0.136	0.129	N/A	0.453
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	0.003	0.000	0.004	0.001	0.001
	1st Quartile	0.007	0.002	0.019	0.002	0.008
	Median	0.020	0.006	0.030	0.003	0.019
	Mean	0.039	0.015	0.039	0.013	0.034
	3rd Quartile	0.040	0.014	0.051	0.008	0.041
	Maximum	0.410	0.204	0.219	0.297	0.235

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $|relB|$ from equation 4.5, with the coefficients' $|c|$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., .032 is 3.2%).

Note: See Table 4.1 for study and model details.

Table F.3 Weighted Distribution of Absolute *T* Values of Empirical Relative Biases by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute *T* Value of the Empirical Bias				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.534	0.466	0.469	2.016	0.823
	1st Quartile	5.249	1.881	11.496	26.018	4.865
	Median	9.393	4.923	39.583	109.619	12.899
	Mean	10.359	7.584	34.061	82.955	11.990
	3rd Quartile	14.051	7.799	53.265	121.253	19.229
	Maximum	27.317	27.197	70.036	187.884	28.692
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	0.244	0.057	0.081	0.216	0.121
	1st Quartile	6.552	2.548	3.304	1.579	0.766
	Median	9.054	5.518	9.561	3.032	2.479
	Mean	11.145	6.938	17.573	3.896	2.900
	3rd Quartile	14.598	9.513	42.500	5.656	4.882
	Maximum	25.982	20.436	46.200	15.265	8.514
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	0.436	0.012	0.010	0.013	0.444
	1st Quartile	1.444	0.317	2.533	0.420	0.861
	Median	2.500	1.008	5.632	2.211	2.532
	Mean	3.032	1.766	7.200	2.808	3.614
	3rd Quartile	3.684	1.930	9.285	4.513	4.982
	Maximum	9.089	9.090	31.114	9.497	17.304
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.191	0.008	0.082	0.234	0.041
	1st Quartile	1.712	0.336	2.047	0.718	1.386
	Median	3.493	1.149	5.566	2.138	2.633
	Mean	3.844	1.435	6.447	2.822	3.036
	3rd Quartile	5.770	2.700	11.040	4.367	4.613
	Maximum	10.506	3.844	13.567	10.755	7.300
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	0.119	0.039	0.157	N/A	0.110
	1st Quartile	2.723	0.628	12.597	N/A	2.707
	Median	4.465	1.274	31.633	N/A	8.701
	Mean	7.633	3.462	32.376	N/A	10.639
	3rd Quartile	8.678	2.254	48.985	N/A	16.767
	Maximum	28.001	22.543	104.669	N/A	38.555
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	0.252	0.078	2.622	0.088	0.030
	1st Quartile	2.198	0.470	7.835	0.720	2.330
	Median	3.946	1.166	11.160	1.567	3.414
	Mean	5.028	1.427	13.969	2.832	4.313
	3rd Quartile	7.401	2.206	15.772	4.233	6.083
	Maximum	12.082	4.885	34.783	10.340	12.434

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across the $|t|$ from equation 4.6, with the coefficients' $|c|$ values used as the weights.

Note: See Table 4.1 for study and model details.

Table F.4 Weighted Distribution of Absolute *T* Values of Empirical Relative Biases by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Absolute *T* Value of the Empirical Bias				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.118	0.167	1.616	5.496	0.377
	1st Quartile	8.347	2.310	16.293	47.025	7.225
	Median	16.759	8.603	67.059	158.048	20.441
	Mean	17.392	13.287	54.761	123.060	19.661
	3rd Quartile	21.822	16.514	86.243	180.315	33.045
	Maximum	46.252	48.484	113.551	288.141	51.849
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	0.053	0.401	0.584	0.016	0.020
	1st Quartile	11.025	4.090	5.743	3.265	1.791
	Median	15.947	10.125	15.378	5.805	3.627
	Mean	18.821	12.005	28.405	5.909	3.535
	3rd Quartile	23.283	18.895	70.007	8.155	4.486
	Maximum	43.190	36.183	70.902	25.824	13.952
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	0.060	0.044	0.003	0.249	0.815
	1st Quartile	1.867	1.163	4.460	0.843	2.043
	Median	4.058	1.883	9.790	3.715	3.919
	Mean	4.315	2.349	11.263	3.988	5.834
	3rd Quartile	5.747	2.589	13.592	5.772	8.206
	Maximum	15.284	8.138	51.358	14.186	25.803
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.320	0.016	0.666	0.373	0.091
	1st Quartile	1.900	1.427	3.577	2.852	1.667
	Median	5.728	2.491	8.431	3.783	5.267
	Mean	6.540	2.861	9.361	4.858	5.171
	3rd Quartile	8.869	3.963	15.513	5.859	8.475
	Maximum	17.381	7.649	20.975	17.675	11.313
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	0.116	0.001	0.367	N/A	1.587
	1st Quartile	5.426	0.730	18.529	N/A	6.402
	Median	7.330	2.295	52.049	N/A	14.152
	Mean	13.750	6.399	51.854	N/A	17.492
	3rd Quartile	16.466	3.247	78.459	N/A	22.536
	Maximum	53.394	42.030	168.476	N/A	58.667
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	0.276	0.262	1.666	0.554	1.213
	1st Quartile	3.617	0.550	11.284	1.318	3.049
	Median	7.605	1.857	17.054	3.286	6.025
	Mean	8.052	2.259	22.402	4.447	6.693
	3rd Quartile	13.194	3.632	25.029	4.135	8.009
	Maximum	19.621	8.810	60.734	18.267	20.083

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across the $|t|$ from equation 4.6, with the coefficients' $|c|$ values used as the weights.

Note: See Table 4.1 for study and model details.

Table F.5 Ratios of Empirical Variance to Full-Sample Variance by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical Variance to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.026	1.027	1.000	1.001	1.000
	1st Quartile	1.045	1.047	1.001	1.002	1.001
	Median	1.050	1.052	1.001	1.003	1.001
	Mean	1.049	1.051	1.006	1.014	1.006
	3rd Quartile	1.053	1.055	1.002	1.006	1.003
	Maximum	1.066	1.068	1.057	1.083	1.057
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.028	1.028	1.003	1.001	1.002
	1st Quartile	1.046	1.049	1.006	1.002	1.004
	Median	1.050	1.053	1.007	1.003	1.005
	Mean	1.049	1.051	1.012	1.005	1.010
	3rd Quartile	1.053	1.055	1.009	1.004	1.006
	Maximum	1.069	1.071	1.070	1.033	1.064
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.032	1.033	1.001	1.000	1.001
	1st Quartile	1.035	1.035	1.001	1.001	1.001
	Median	1.041	1.041	1.002	1.001	1.002
	Mean	1.040	1.041	1.010	1.006	1.011
	3rd Quartile	1.044	1.045	1.018	1.008	1.015
	Maximum	1.049	1.050	1.039	1.023	1.043
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	1.034	1.033	1.001	1.001	1.001
	1st Quartile	1.036	1.036	1.002	1.001	1.002
	Median	1.040	1.040	1.003	1.002	1.003
	Mean	1.042	1.042	1.011	1.006	1.012
	3rd Quartile	1.045	1.045	1.014	1.007	1.013
	Maximum	1.066	1.067	1.049	1.027	1.062
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.038	1.038	1.001	N/A	1.001
	1st Quartile	1.040	1.042	1.002	N/A	1.002
	Median	1.046	1.047	1.019	N/A	1.021
	Mean	1.047	1.048	1.019	N/A	1.020
	3rd Quartile	1.052	1.054	1.030	N/A	1.031
	Maximum	1.074	1.075	1.048	N/A	1.047
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.034	1.035	1.002	1.001	1.001
	1st Quartile	1.038	1.039	1.003	1.002	1.003
	Median	1.045	1.046	1.015	1.008	1.014
	Mean	1.046	1.047	1.018	1.011	1.018
	3rd Quartile	1.053	1.053	1.024	1.014	1.029
	Maximum	1.070	1.070	1.051	1.033	1.046

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of v / v_f from equation 4.7, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.6 Ratios of Empirical Variance to Full-Sample Variance by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical Variance to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.068	1.076	1.001	1.002	1.001
	1st Quartile	1.121	1.136	1.002	1.005	1.002
	Median	1.129	1.144	1.003	1.006	1.003
	Mean	1.129	1.144	1.013	1.029	1.014
	3rd Quartile	1.139	1.158	1.006	1.011	1.006
	Maximum	1.189	1.200	1.138	1.178	1.140
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.075	1.082	1.007	1.003	1.006
	1st Quartile	1.121	1.133	1.014	1.006	1.011
	Median	1.131	1.147	1.017	1.007	1.013
	Mean	1.130	1.143	1.029	1.013	1.024
	3rd Quartile	1.145	1.156	1.021	1.009	1.019
	Maximum	1.178	1.187	1.151	1.084	1.157
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.084	1.086	1.002	1.001	1.002
	1st Quartile	1.101	1.102	1.003	1.001	1.003
	Median	1.109	1.113	1.005	1.003	1.005
	Mean	1.112	1.114	1.026	1.015	1.026
	3rd Quartile	1.122	1.122	1.044	1.021	1.039
	Maximum	1.145	1.146	1.108	1.063	1.112
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	1.085	1.087	1.002	1.001	1.002
	1st Quartile	1.098	1.101	1.005	1.003	1.004
	Median	1.107	1.109	1.007	1.005	1.007
	Mean	1.115	1.118	1.026	1.016	1.027
	3rd Quartile	1.122	1.125	1.035	1.017	1.032
	Maximum	1.185	1.198	1.128	1.073	1.155
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.098	1.102	1.002	N/A	1.002
	1st Quartile	1.115	1.123	1.006	N/A	1.006
	Median	1.127	1.136	1.049	N/A	1.045
	Mean	1.130	1.138	1.049	N/A	1.052
	3rd Quartile	1.143	1.153	1.078	N/A	1.085
	Maximum	1.203	1.211	1.108	N/A	1.125
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.090	1.093	1.004	1.002	1.003
	1st Quartile	1.098	1.102	1.007	1.004	1.008
	Median	1.119	1.125	1.037	1.021	1.035
	Mean	1.124	1.132	1.042	1.026	1.043
	3rd Quartile	1.140	1.147	1.057	1.034	1.063
	Maximum	1.185	1.190	1.120	1.084	1.121

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of v / v_f from equation 4.7, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.7 Ratios of Empirical MSE to Full-Sample Variance by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical MSE to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.029	1.033	1.000	1.002	1.000
	1st Quartile	1.050	1.050	1.001	1.003	1.001
	Median	1.054	1.054	1.002	1.005	1.001
	Mean	1.054	1.054	1.008	1.105	1.008
	3rd Quartile	1.056	1.059	1.007	1.062	1.005
	Maximum	1.073	1.075	1.059	1.846	1.068
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.037	1.033	1.003	1.001	1.002
	1st Quartile	1.048	1.049	1.006	1.002	1.004
	Median	1.051	1.053	1.007	1.003	1.005
	Mean	1.052	1.053	1.014	1.005	1.010
	3rd Quartile	1.055	1.056	1.011	1.004	1.007
	Maximum	1.082	1.077	1.071	1.033	1.066
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.032	1.033	1.001	1.000	1.001
	1st Quartile	1.035	1.035	1.001	1.001	1.001
	Median	1.041	1.041	1.002	1.001	1.002
	Mean	1.041	1.041	1.011	1.006	1.011
	3rd Quartile	1.045	1.045	1.018	1.008	1.016
	Maximum	1.049	1.050	1.045	1.023	1.047
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	1.034	1.033	1.001	1.001	1.001
	1st Quartile	1.038	1.037	1.002	1.001	1.002
	Median	1.040	1.040	1.003	1.002	1.003
	Mean	1.043	1.043	1.011	1.006	1.013
	3rd Quartile	1.047	1.046	1.015	1.007	1.014
	Maximum	1.066	1.068	1.052	1.027	1.065
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.038	1.039	1.001	N/A	1.001
	1st Quartile	1.041	1.042	1.014	N/A	1.008
	Median	1.048	1.050	1.026	N/A	1.021
	Mean	1.051	1.049	1.026	N/A	1.024
	3rd Quartile	1.056	1.056	1.039	N/A	1.035
	Maximum	1.111	1.087	1.079	N/A	1.073
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.034	1.035	1.002	1.001	1.001
	1st Quartile	1.038	1.039	1.004	1.002	1.003
	Median	1.046	1.046	1.015	1.008	1.014
	Mean	1.047	1.047	1.019	1.011	1.019
	3rd Quartile	1.054	1.054	1.025	1.014	1.030
	Maximum	1.071	1.071	1.052	1.035	1.050

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of m / v_f from equation 4.8, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.8 Ratios of Empirical MSE to Full-Sample Variance by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Empirical MSE to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.088	1.107	1.001	1.004	1.001
	1st Quartile	1.143	1.145	1.002	1.006	1.002
	Median	1.160	1.159	1.006	1.014	1.003
	Mean	1.167	1.172	1.025	1.420	1.025
	3rd Quartile	1.183	1.183	1.032	1.227	1.019
	Maximum	1.322	1.364	1.174	4.281	1.169
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.118	1.124	1.007	1.003	1.007
	1st Quartile	1.139	1.141	1.015	1.006	1.012
	Median	1.150	1.153	1.017	1.007	1.013
	Mean	1.157	1.157	1.041	1.014	1.027
	3rd Quartile	1.162	1.166	1.040	1.010	1.023
	Maximum	1.353	1.304	1.219	1.092	1.176
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.086	1.086	1.002	1.001	1.002
	1st Quartile	1.103	1.103	1.003	1.002	1.003
	Median	1.111	1.113	1.006	1.003	1.006
	Mean	1.114	1.115	1.030	1.016	1.031
	3rd Quartile	1.122	1.123	1.045	1.022	1.040
	Maximum	1.150	1.150	1.138	1.066	1.150
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	1.087	1.088	1.002	1.001	1.002
	1st Quartile	1.106	1.101	1.005	1.003	1.005
	Median	1.111	1.110	1.007	1.005	1.007
	Mean	1.120	1.119	1.028	1.017	1.030
	3rd Quartile	1.130	1.126	1.038	1.017	1.037
	Maximum	1.193	1.202	1.141	1.074	1.176
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.098	1.105	1.003	N/A	1.003
	1st Quartile	1.117	1.123	1.040	N/A	1.033
	Median	1.133	1.137	1.080	N/A	1.058
	Mean	1.164	1.153	1.092	N/A	1.083
	3rd Quartile	1.167	1.163	1.109	N/A	1.116
	Maximum	1.564	1.350	1.301	N/A	1.328
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.090	1.093	1.004	1.002	1.003
	1st Quartile	1.098	1.102	1.009	1.004	1.008
	Median	1.125	1.128	1.040	1.021	1.039
	Mean	1.132	1.133	1.050	1.027	1.048
	3rd Quartile	1.152	1.151	1.073	1.035	1.066
	Maximum	1.202	1.193	1.144	1.102	1.141

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of m / v_f from equation 4.8, with the coefficients' $\sqrt{v_f}$ values used as the weights. To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.9 Empirical Relative Biases of Variance Estimates by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of Variance Estimates: LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	-0.022	-0.011	-0.063	-0.183	-0.059
	1st Quartile	-0.003	-0.003	-0.004	-0.012	-0.003
	Median	0.000	0.000	-0.003	-0.004	-0.003
	Mean	-0.002	-0.001	-0.007	-0.026	-0.007
	3rd Quartile	0.003	0.001	-0.001	-0.002	-0.001
	Maximum	0.010	0.006	0.000	0.001	0.000
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	-0.016	-0.016	-0.075	-0.007	-0.061
	1st Quartile	-0.007	-0.006	-0.014	-0.002	-0.006
	Median	-0.003	-0.003	-0.012	-0.001	-0.004
	Mean	-0.003	-0.004	-0.017	-0.002	-0.009
	3rd Quartile	0.001	-0.002	-0.010	-0.001	-0.003
	Maximum	0.011	0.002	-0.004	0.001	0.001
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	-0.010	-0.009	-0.054	-0.010	-0.059
	1st Quartile	-0.003	-0.003	-0.024	-0.003	-0.019
	Median	0.000	-0.001	-0.004	-0.001	-0.004
	Mean	0.000	-0.002	-0.014	-0.002	-0.014
	3rd Quartile	0.002	0.000	-0.002	0.000	-0.002
	Maximum	0.012	0.008	0.000	0.001	0.001
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	-0.061	-0.037	-0.061	-0.025	-0.075
	1st Quartile	-0.013	-0.010	-0.042	-0.004	-0.045
	Median	-0.006	-0.004	-0.005	-0.001	-0.006
	Mean	-0.010	-0.008	-0.016	-0.004	-0.018
	3rd Quartile	-0.002	-0.001	-0.004	-0.001	-0.003
	Maximum	0.011	0.004	0.003	0.001	0.003
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	-0.026	-0.007	-0.068	N/A	-0.065
	1st Quartile	-0.012	-0.003	-0.042	N/A	-0.042
	Median	-0.010	-0.001	-0.029	N/A	-0.029
	Mean	-0.010	-0.001	-0.030	N/A	-0.028
	3rd Quartile	-0.005	0.000	-0.006	N/A	-0.004
	Maximum	-0.002	0.003	-0.003	N/A	-0.001
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	-0.021	-0.017	-0.062	-0.022	-0.062
	1st Quartile	-0.013	-0.005	-0.037	-0.010	-0.053
	Median	-0.007	-0.004	-0.021	-0.004	-0.020
	Mean	-0.009	-0.005	-0.025	-0.006	-0.027
	3rd Quartile	-0.004	-0.002	-0.005	-0.001	-0.004
	Maximum	0.003	0.003	-0.002	0.001	-0.001

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $(e-v)/v$ from equations 4.7 and 4.9, with the coefficients' $\sqrt{v}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table F.10 Empirical Relative Biases of Variance Estimates by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of Variance Estimates				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	-0.052	-0.029	-0.141	-0.335	-0.125
	1st Quartile	-0.001	-0.007	-0.010	-0.026	-0.007
	Median	0.006	-0.001	-0.007	-0.008	-0.006
	Mean	0.002	-0.003	-0.016	-0.050	-0.015
	3rd Quartile	0.008	0.005	-0.003	-0.004	-0.003
	Maximum	0.021	0.010	0.000	0.003	-0.001
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	-0.032	-0.038	-0.153	-0.016	-0.137
	1st Quartile	-0.014	-0.016	-0.033	-0.005	-0.016
	Median	-0.003	-0.012	-0.029	-0.003	-0.014
	Mean	-0.004	-0.012	-0.039	-0.004	-0.023
	3rd Quartile	0.002	-0.007	-0.025	-0.002	-0.010
	Maximum	0.031	0.004	-0.009	0.002	0.000
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	-0.025	-0.023	-0.123	-0.029	-0.130
	1st Quartile	-0.010	-0.011	-0.057	-0.008	-0.044
	Median	-0.004	-0.008	-0.010	-0.003	-0.011
	Mean	-0.004	-0.008	-0.034	-0.006	-0.035
	3rd Quartile	0.000	-0.004	-0.004	-0.001	-0.006
	Maximum	0.017	0.007	0.001	0.003	0.001
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	-0.134	-0.098	-0.146	-0.064	-0.156
	1st Quartile	-0.027	-0.028	-0.096	-0.013	-0.089
	Median	-0.016	-0.012	-0.013	-0.005	-0.014
	Mean	-0.023	-0.022	-0.038	-0.009	-0.039
	3rd Quartile	-0.003	-0.005	-0.007	-0.003	-0.007
	Maximum	0.024	0.012	0.006	0.002	0.006
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	-0.067	-0.021	-0.153	N/A	-0.163
	1st Quartile	-0.035	-0.009	-0.103	N/A	-0.114
	Median	-0.021	-0.006	-0.079	N/A	-0.069
	Mean	-0.024	-0.007	-0.072	N/A	-0.073
	3rd Quartile	-0.009	-0.003	-0.015	N/A	-0.010
	Maximum	-0.002	0.003	-0.007	N/A	-0.003
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	-0.048	-0.041	-0.136	-0.049	-0.134
	1st Quartile	-0.028	-0.015	-0.082	-0.019	-0.107
	Median	-0.019	-0.007	-0.054	-0.009	-0.046
	Mean	-0.019	-0.011	-0.058	-0.014	-0.061
	3rd Quartile	-0.006	-0.003	-0.014	-0.003	-0.009
	Maximum	0.011	0.009	-0.005	0.002	-0.004

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $(e - v) / v$ from equations 4.7 and 4.9, with the coefficients' $\sqrt{v}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table F.11 Empirical Relative Biases of MSE Estimates by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of MSE Estimates				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	-0.026	-0.024	-0.065	-0.521	-0.066
	1st Quartile	-0.011	-0.008	-0.009	-0.067	-0.005
	Median	-0.005	-0.003	-0.003	-0.007	-0.003
	Mean	-0.006	-0.004	-0.009	-0.089	-0.008
	3rd Quartile	0.000	0.000	-0.001	-0.002	-0.001
	Maximum	0.006	0.006	-0.001	0.001	0.000
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	-0.026	-0.019	-0.076	-0.008	-0.063
	1st Quartile	-0.010	-0.008	-0.016	-0.002	-0.006
	Median	-0.007	-0.005	-0.013	-0.001	-0.005
	Mean	-0.006	-0.006	-0.019	-0.002	-0.009
	3rd Quartile	-0.002	-0.003	-0.010	-0.001	-0.003
	Maximum	0.011	0.001	-0.004	0.001	0.001
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	-0.010	-0.009	-0.060	-0.010	-0.067
	1st Quartile	-0.003	-0.003	-0.024	-0.003	-0.020
	Median	-0.001	-0.001	-0.004	-0.001	-0.004
	Mean	-0.001	-0.002	-0.014	-0.003	-0.015
	3rd Quartile	0.001	0.000	-0.002	0.000	-0.002
	Maximum	0.009	0.006	0.000	0.001	0.001
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	-0.061	-0.037	-0.061	-0.025	-0.076
	1st Quartile	-0.014	-0.010	-0.042	-0.004	-0.046
	Median	-0.007	-0.004	-0.005	-0.002	-0.006
	Mean	-0.011	-0.008	-0.016	-0.004	-0.018
	3rd Quartile	-0.003	-0.001	-0.004	-0.001	-0.003
	Maximum	0.011	0.004	0.003	0.001	0.002
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	-0.045	-0.018	-0.096	N/A	-0.089
	1st Quartile	-0.014	-0.004	-0.046	N/A	-0.044
	Median	-0.011	-0.002	-0.035	N/A	-0.030
	Mean	-0.013	-0.003	-0.036	N/A	-0.032
	3rd Quartile	-0.007	0.000	-0.018	N/A	-0.008
	Maximum	-0.002	0.003	-0.003	N/A	-0.002
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	-0.025	-0.017	-0.068	-0.024	-0.066
	1st Quartile	-0.013	-0.005	-0.037	-0.011	-0.054
	Median	-0.008	-0.004	-0.023	-0.004	-0.021
	Mean	-0.010	-0.005	-0.027	-0.007	-0.028
	3rd Quartile	-0.004	-0.002	-0.005	-0.001	-0.005
	Maximum	0.001	0.003	-0.002	0.001	-0.001

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $\left(e-m\right)/m$ from equations 4.8 and 4.9, with the coefficients' $\sqrt{m}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table F.12 Empirical Relative Biases of MSE Estimates by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Empirical Relative Bias of MSE Estimates				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	-0.135	-0.154	-0.147	-0.821	-0.147
	1st Quartile	-0.047	-0.031	-0.038	-0.235	-0.020
	Median	-0.018	-0.014	-0.008	-0.025	-0.006
	Mean	-0.031	-0.026	-0.027	-0.218	-0.025
	3rd Quartile	-0.003	-0.001	-0.004	-0.006	-0.004
	Maximum	0.009	0.007	-0.002	0.003	-0.001
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	-0.150	-0.113	-0.195	-0.026	-0.149
	1st Quartile	-0.037	-0.023	-0.048	-0.005	-0.019
	Median	-0.022	-0.017	-0.031	-0.004	-0.014
	Mean	-0.027	-0.024	-0.049	-0.005	-0.025
	3rd Quartile	-0.014	-0.014	-0.025	-0.003	-0.010
	Maximum	0.028	0.004	-0.009	0.002	0.000
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	-0.029	-0.024	-0.156	-0.029	-0.167
	1st Quartile	-0.013	-0.013	-0.058	-0.008	-0.054
	Median	-0.005	-0.008	-0.011	-0.003	-0.012
	Mean	-0.006	-0.009	-0.037	-0.007	-0.038
	3rd Quartile	-0.001	-0.004	-0.005	-0.001	-0.006
	Maximum	0.016	0.006	0.001	0.003	0.001
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	-0.137	-0.100	-0.146	-0.066	-0.156
	1st Quartile	-0.034	-0.028	-0.099	-0.014	-0.094
	Median	-0.018	-0.013	-0.013	-0.005	-0.015
	Mean	-0.027	-0.022	-0.040	-0.010	-0.041
	3rd Quartile	-0.003	-0.006	-0.007	-0.003	-0.008
	Maximum	0.023	0.012	0.006	0.002	0.006
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	-0.246	-0.116	-0.277	N/A	-0.288
	1st Quartile	-0.055	-0.013	-0.154	N/A	-0.147
	Median	-0.035	-0.008	-0.110	N/A	-0.074
	Mean	-0.051	-0.019	-0.108	N/A	-0.098
	3rd Quartile	-0.019	-0.004	-0.054	N/A	-0.040
	Maximum	-0.003	0.001	-0.009	N/A	-0.003
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	-0.074	-0.044	-0.159	-0.064	-0.164
	1st Quartile	-0.029	-0.016	-0.084	-0.019	-0.109
	Median	-0.024	-0.007	-0.060	-0.009	-0.049
	Mean	-0.025	-0.012	-0.065	-0.015	-0.065
	3rd Quartile	-0.011	-0.005	-0.015	-0.003	-0.009
	Maximum	0.005	0.007	-0.005	0.001	-0.004

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; MSE = mean squared error; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: Weighted distributions are across coefficients of $(e-m)/m$ from equations 4.8 and 4.9, with the coefficients' $\sqrt{m}$ values used as the weights. Relative biases are in fractional form. Multiply each by 100 percent to put it in percentage form (e.g., -.022 = -2.2%).

Note: See Table 4.1 for study and model details.

Table F.13 Ratios of Estimated Variance to Full-Sample Variance by MIVRA Method for 5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Estimated Variance to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.023	1.024	0.987	0.885	0.994
	1st Quartile	1.042	1.047	0.998	0.992	0.999
	Median	1.049	1.052	0.999	0.999	0.999
	Mean	1.047	1.049	0.999	0.988	0.999
	3rd Quartile	1.053	1.055	0.999	1.001	1.000
	Maximum	1.069	1.066	1.005	1.004	1.003
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.028	1.028	0.989	1.000	0.997
	1st Quartile	1.041	1.043	0.993	1.001	1.000
	Median	1.045	1.046	0.995	1.002	1.001
	Mean	1.046	1.046	0.995	1.004	1.001
	3rd Quartile	1.052	1.052	0.996	1.003	1.002
	Maximum	1.068	1.063	1.002	1.025	1.003
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.031	1.031	0.978	1.000	0.977
	1st Quartile	1.036	1.035	0.992	1.000	0.995
	Median	1.039	1.039	0.998	1.001	0.999
	Mean	1.040	1.039	0.996	1.003	0.996
	3rd Quartile	1.042	1.042	0.999	1.005	1.000
	Maximum	1.055	1.051	1.002	1.014	1.002
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.990	1.017	0.965	0.993	0.955
	1st Quartile	1.029	1.030	0.990	0.999	0.990
	Median	1.033	1.035	0.997	1.001	0.997
	Mean	1.032	1.034	0.994	1.003	0.994
	3rd Quartile	1.040	1.037	1.000	1.002	0.999
	Maximum	1.056	1.054	1.004	1.025	1.008
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.025	1.037	0.975	N/A	0.977
	1st Quartile	1.030	1.043	0.986	N/A	0.988
	Median	1.035	1.046	0.989	N/A	0.991
	Mean	1.037	1.047	0.989	N/A	0.991
	3rd Quartile	1.044	1.050	0.996	N/A	0.998
	Maximum	1.061	1.068	0.998	N/A	1.000
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.020	1.032	0.978	0.998	0.973
	1st Quartile	1.026	1.035	0.987	1.001	0.980
	Median	1.037	1.044	0.994	1.003	0.995
	Mean	1.037	1.042	0.992	1.004	0.990
	3rd Quartile	1.044	1.048	0.998	1.005	0.998
	Maximum	1.054	1.053	1.000	1.011	1.002

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.14 Ratios of Estimated Variance to Full-Sample Variance by MIVRA Method for 12.5 Percent Deletion Rate

Study/Model	Summary Statistics	Ratio of Estimated Variance to Full-Sample Variance				
		LD	REWT	WSHD	MPLUS	CTBHD
N4/SPDMON	Sample Size	32	32	32	32	32
	Minimum	1.063	1.067	0.968	0.767	0.984
	1st Quartile	1.113	1.136	0.995	0.985	0.997
	Median	1.135	1.146	0.998	0.998	0.998
	Mean	1.131	1.141	0.996	0.975	0.998
	3rd Quartile	1.147	1.157	0.999	1.002	0.999
	Maximum	1.199	1.191	1.012	1.008	1.008
N4/MHTRT	Sample Size	34	34	34	34	34
	Minimum	1.077	1.080	0.975	1.001	0.994
	1st Quartile	1.110	1.120	0.984	1.003	0.998
	Median	1.124	1.126	0.987	1.005	1.000
	Mean	1.126	1.129	0.988	1.009	1.000
	3rd Quartile	1.141	1.145	0.991	1.008	1.002
	Maximum	1.190	1.178	1.006	1.067	1.010
N14/YOTMTHLP	Sample Size	35	35	35	35	35
	Minimum	1.080	1.082	0.946	0.998	0.944
	1st Quartile	1.093	1.093	0.981	1.000	0.980
	Median	1.104	1.104	0.995	1.001	0.996
	Mean	1.107	1.105	0.990	1.009	0.990
	3rd Quartile	1.114	1.114	0.999	1.013	0.998
	Maximum	1.151	1.136	1.004	1.037	1.004
N14/YORXHLP	Sample Size	34	34	34	34	34
	Minimum	0.979	1.048	0.913	0.985	0.907
	1st Quartile	1.080	1.082	0.977	0.998	0.976
	Median	1.092	1.096	0.993	1.001	0.993
	Mean	1.089	1.094	0.987	1.007	0.986
	3rd Quartile	1.108	1.102	0.999	1.005	0.999
	Maximum	1.155	1.150	1.008	1.064	1.008
N19/ANLYR	Sample Size	26	26	26	N/A	26
	Minimum	1.066	1.104	0.938	N/A	0.932
	1st Quartile	1.078	1.117	0.961	N/A	0.957
	Median	1.099	1.130	0.969	N/A	0.976
	Mean	1.103	1.130	0.972	N/A	0.974
	3rd Quartile	1.123	1.140	0.991	N/A	0.995
	Maximum	1.179	1.196	0.996	N/A	1.000
N19/ABODANL	Sample Size	26	26	26	26	26
	Minimum	1.056	1.092	0.949	0.995	0.948
	1st Quartile	1.071	1.097	0.971	1.003	0.951
	Median	1.113	1.121	0.985	1.012	0.988
	Mean	1.103	1.119	0.981	1.011	0.979
	3rd Quartile	1.121	1.136	0.994	1.017	0.998
	Maximum	1.151	1.154	1.000	1.031	1.004

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: To convert a ratio to a percentage difference, one would subtract 1, then multiply by 100. For example, a ratio of 1.159 represents a percentage increase of 15.9 percent.

Note: See Table 4.1 for study and model details.

Table F.15 Statistical Significance of Covariates by MIVRA Method, Simulation versus Full Sample: 5 Percent Deletion Rate

	LD		REWT		WSHD		MPLUS		CTBHD	
Study/Model	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**
N4/SPDMON										
α = 0.05 (23 coefficients)	71.4	0.5	71.6	0.3	71.9	0.0	71.9	0.0	71.9	0.0
N.S. (9 coefficients)	1.1	27.0	1.4	26.7	0.0	28.1	0.1	28.0	0.0	28.1
N4/MHTRT										
α = 0.05 (12 coefficients)	32.9	2.3	33.0	2.3	35.2	0.1	35.3	0.0	35.2	0.1
N.S. (22 coefficients)	2.1	62.6	2.0	62.7	1.7	63.0	1.0	63.7	1.3	63.4
N14/YOTMTHLP										
α = 0.05 (13 coefficients)	32.9	4.2	33.0	4.2	36.5	0.7	36.6	0.5	36.5	0.7
N.S. (22 coefficients)	1.2	61.7	1.2	61.6	0.4	62.5	0.2	62.6	0.5	62.4
N14/YORXHLP										
α = 0.05 (9 coefficients)	22.9	3.6	22.8	3.6	24.5	2.0	24.9	1.6	24.5	1.9
N.S. (25 coefficients)	2.1	71.4	2.2	71.3	0.8	72.8	0.5	73.1	0.8	72.7
N19/ANLYR										
α = 0.05 (21 coefficients)	79.0	1.8	78.6	2.1	79.6	1.2	N/A	N/A	79.4	1.3
N.S. (5 coefficients)	1.1	18.2	1.1	18.2	0.0	19.2	N/A	N/A	0.0	19.2
N19/ABODANL										
α = 0.05 (11 coefficients)	39.3	3.1	38.8	3.6	40.7	1.6	41.7	0.6	40.9	1.4
N.S. (15 coefficients)	1.8	55.9	1.5	56.2	1.3	56.4	1.2	56.5	1.3	56.4

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: See Table 4.1 for study and model details.

Table F.16 Statistical Significance of Covariates by MIVRA Method, Simulation versus Full Sample: 12.5 Percent Deletion Rate

	LD		REWT		WSHD		MPLUS		CTBHD	
Study/Model	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**	**α = 0.05**	**N.S.**
N4/SPDMON										
α = 0.05 (23 coefficients)	70.6	1.3	71.2	0.7	71.9	0.0	71.9	0.0	71.9	0.0
N.S. (9 coefficients)	1.5	26.6	2.1	26.0	0.1	28.0	0.3	27.8	0.1	28.1
N4/MHTRT										
α = 0.05 (12 coefficients)	30.5	4.8	30.4	4.9	34.7	0.6	35.2	0.1	34.8	0.5
N.S. (22 coefficients)	3.2	61.6	2.9	61.8	2.7	62.0	1.4	63.3	1.9	62.8
N14/YOTMTHLP										
α = 0.05 (13 coefficients)	31.0	6.2	31.1	6.1	35.9	1.3	36.2	0.9	35.7	1.4
N.S. (22 coefficients)	1.8	61.1	1.9	60.9	1.1	61.8	0.5	62.4	1.1	61.8
N14/YORXHLP										
α = 0.05 (9 coefficients)	21.4	5.1	21.3	5.2	23.5	3.0	24.0	2.5	23.4	3.1
N.S. (25 coefficients)	3.0	70.6	3.2	70.3	1.6	71.9	1.0	72.5	1.5	72.1
N19/ANLYR										
α = 0.05 (21 coefficients)	77.2	3.6	76.5	4.3	79.5	1.3	N/A	N/A	79.4	1.4
N.S. (5 coefficients)	1.5	17.7	1.6	17.6	0.3	18.9	N/A	N/A	0.0	19.2
N19/ABODANL										
α = 0.05 (11 coefficients)	38.2	4.1	37.0	5.3	38.9	3.4	40.8	1.5	38.9	3.4
N.S. (15 coefficients)	2.1	55.6	1.6	56.1	1.5	56.2	1.5	56.2	1.4	56.3

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. For the N19/ANLYR model, Mplus failed to run, reporting that the Fisher information matrix was ill-conditioned and non-positive definite.

Note: See Table 4.1 for study and model details.

Table F.17 Statistical Significance of Individual Covariates in N4/SPDMON by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD		LG	
		$\alpha = 0.05$	N.S.	$\alpha = 0.05$	N.S.	$\alpha = 0.05$	N.S.	$\alpha = 0.05$	N.S.	$\alpha = 0.05$	N.S.	$\alpha = 0.05$	N.S.
Pregnancy Status													
Pregnant	0.0002	1,600	0	1,599	1	1,600	0	1,600	0	400	0	49	0
Postpartum	0.0661	385	1,215	526	1,074	65	1,535	270	1,330	10	390	0	49
Not Pregnant or Postpartum	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age Group													
18–25	0.0015	1,599	1	1,583	17	1,600	0	1,600	0	400	0	49	0
26–34	0.0060	1,587	13	1,474	126	1,600	0	1,600	0	400	0	49	0
35–44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity													
White	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Black/African American	0.0022	1,516	84	1,537	63	1,600	0	1,600	0	400	0	49	0
Other	0.0226	693	907	1,218	382	1,600	0	1,600	0	400	0	49	0
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Marital Status													
Married	0.0999	55	1,545	140	1,460	55	1,545	53	1,547	7	393	0	49
Widowed/Divorced/ Separated	0.0005	1,597	3	1,597	3	1,600	0	1,600	0	400	0	49	0
Never Married	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Education													
Less than High School	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
High School	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Some College	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
College Graduate	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Employment Status													
Employed Full Time	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Employed Part Time	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Unemployed	0.5674	0	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49
Other	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Federal Poverty Level (FPL)													
Below FPL	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
100%–199% FPL	0.0019	1,583	17	1,574	26	1,587	13	1,600	0	398	2	49	0
≥ 200% FPL	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.17 Statistical Significance of Individual Covariates in N4/SPDMON by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD		LG	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Rapid Repeat Birth													
Current	0.0756	169	1,431	365	1,235	0	1,600	8	1,592	0	400	0	49
Past	0.6635	3	1,597	4	1,596	0	1,600	0	1,600	0	400	0	49
Neither	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Number of Biological Children in Household													
0	0.1909	6	1,594	25	1,575	0	1,600	0	1,600	0	400	0	49
1	0.9899	0	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49
2	0.1989	170	1,430	109	1,491	0	1,600	0	1,600	0	400	0	49
$\geq$ 3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Insurance													
Insured	0.5998	1	1,599	5	1,595	0	1,600	0	1,600	0	400	0	49
Uninsured	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Status													
Excellent	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Very Good	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Good	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Fair/Poor	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Problems													
0	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
1	0.0000	1,600	0	1,600	0	1,600	0	1,549	51	400	0	49	0
$\geq$ 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Month Cigarette Use													
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Alcohol Use Disorder													
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.17 Statistical Significance of Individual Covariates in N4/SPDMON by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD		LG	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Past Year Illicit Drug Use Disorder													
No Illicit Drug Use	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Illicit Drug Use, No Substance Use Disorder	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0	49	0
Illicit Drug Use, Substance Use Disorder	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; LG = Latent GOLD; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. These rows correspond to reference levels of covariates.

Note: See Table 4.1 for study and model details.

Table F.18 Statistical Significance of Individual Covariates in N4/MHTRT by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Pregnancy Status											
Pregnant	0.1942	62	1,538	111	1,489	0	1,600	0	1,600	1	399
Postpartum	0.2718	74	1,526	35	1,565	0	1,600	0	1,600	0	400
Not Pregnant or Postpartum	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age Group											
18–25	0.0001	1,600	0	1,600	0	1,600	0	1,600	0	400	0
26–34	0.0582	477	1,123	431	1,169	352	1,248	330	1,270	112	288
35–44	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity											
White	0.0001	1,600	0	1,600	0	1,600	0	1,600	0	400	0
Black/African American	0.1937	6	1,594	35	1,565	0	1,600	0	1,600	0	400
Other	0.1932	121	1,479	95	1,505	0	1,600	0	1,600	0	400
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Marital Status											
Married	0.5642	3	1,597	2	1,598	0	1,600	0	1,600	0	400
Widowed/Divorced/Separated	0.5300	1	1,599	2	1,598	0	1,600	0	1,600	0	400
Never Married	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Education											
Less than High School	0.0008	1,530	70	1,567	33	1,600	0	1,600	0	400	0
High School	0.0002	1,593	7	1,594	6	1,600	0	1,600	0	400	0
Some College	0.0342	146	1,454	244	1,356	1,452	148	1,548	52	357	43
College Graduate	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Employment Status											
Employed Full Time	0.1847	735	865	624	976	9	1,591	0	1,600	0	400
Employed Part Time	0.0303	1,196	404	1,172	428	1,574	26	1,551	49	386	14
Unemployed	0.5220	15	1,585	34	1,566	0	1,600	0	1,600	0	400
Other	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Federal Poverty Level (FPL)											
Below FPL	0.3630	8	1,592	23	1,577	15	1,585	1	1,599	1	399
100%–199% FPL	0.5356	0	1,600	2	1,598	0	1,600	0	1,600	0	400
≥ 200% FPL	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rapid Repeat Birth											
Current	0.6736	0	1,600	3	1,597	0	1,600	0	1,600	0	400
Past	0.5740	3	1,597	2	1,598	0	1,600	0	1,600	0	400
Neither	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.18 Statistical Significance of Individual Covariates in N4/MHTRT by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Number of Biological Children in Household											
0	0.9838	0	1,600	0	1,600	0	1,600	0	1,600	0	400
1	0.5405	0	1,600	2	1,598	0	1,600	0	1,600	0	400
2	0.9870	0	1,600	0	1,600	0	1,600	0	1,600	0	400
≥ 3	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Insurance											
Insured	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
Uninsured	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Status											
Excellent	0.3491	12	1,588	25	1,575	2	1,598	0	1,600	0	400
Very Good	0.0851	551	1,049	387	1,213	779	821	33	1,567	57	343
Good	0.0314	1,210	390	976	624	1,571	29	1,530	70	376	24
Fair/Poor	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Problems											
0	0.0118	1,032	568	1,027	573	1,309	291	1,587	13	380	20
1	0.2120	43	1,557	69	1,531	4	1,596	0	1,600	1	399
≥ 2	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Month Cigarette Use											
Yes	0.2414	4	1,596	5	1,595	0	1,600	0	1,600	0	400
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Alcohol Use Disorder											
Yes	0.3639	29	1,571	17	1,583	0	1,600	0	1,600	0	400
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Illicit Drug Use Disorder											
No Illicit Drug Use	0.0585	259	1,341	320	1,280	803	797	554	1,046	141	259
Illicit Drug Use, No Substance Use Disorder	0.0126	982	618	969	631	1,534	66	1,583	17	379	21
Illicit Drug Use, Substance Use Disorder	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.18 Statistical Significance of Individual Covariates in N4/MHTRT by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Had Depression in Lifetime											
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Had Anxiety in Lifetime											
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. These rows correspond to reference levels of covariates.

Note: See Table 4.1 for study and model details.

Table F.19 Statistical Significance of Individual Covariates in N14/YOTMTHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Age											
12	0.0412	801	799	827	773	1,560	40	1,575	25	385	15
13	0.7216	0	1,600	1	1,599	0	1,600	0	1,600	0	400
14	0.5449	1	1,599	1	1,599	0	1,600	0	1,600	0	400
15	0.7212	0	1,600	0	1,600	0	1,600	0	1,600	0	400
16	0.7180	0	1,600	0	1,600	0	1,600	0	1,600	0	400
17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Gender											
Male	0.9506	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity											
White	0.0128	1,389	211	1,394	206	1,600	0	1,600	0	400	0
Black/African American	0.4798	3	1,597	3	1,597	0	1,600	0	1,600	0	400
Other	0.2291	56	1,544	52	1,548	0	1,600	0	1,600	0	400
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Income											
Less than $20,000	0.0657	220	1,380	309	1,291	387	1,213	86	1,514	95	305
$20,000–$49,999	0.3798	3	1,597	4	1,596	0	1,600	0	1,600	0	400
$50,000–$74,999	0.2075	13	1,587	29	1,571	0	1,600	0	1,600	0	400
$75,000–$99,999	0.0478	464	1,136	599	1,001	1,257	343	1,141	459	288	112
$100,000 or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Insurance											
Has Insurance	0.1003	231	1,369	249	1,351	0	1,600	0	1,600	0	400
Does Not Have Insurance	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rural/Urban Segment											
Rural	0.9195	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Urban	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Number of Delinquent Behaviors											
None	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
One	0.0005	1,599	1	1,598	2	1,600	0	1,600	0	400	0
Two or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.19 Statistical Significance of Individual Covariates in N14/YOTMTHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	**α** = 0.05	N.S.	**α** = 0.05	N.S.	**α** = 0.05	N.S.	**α** = 0.05	N.S.
Grade for Last Semester or Grading Period Completed											
An "A+," "A," or "A-Minus" Average	0.0755	359	1,241	298	1,302	315	1,285	213	1,387	70	330
A "B+," "B," or "B-Minus" Average	0.0117	1,209	391	1,232	368	1,279	321	1,422	178	315	85
A "C+," "C," or "C-Minus" Average	0.2402	21	1,579	15	1,585	13	1,587	2	1,598	4	396
A "D" or Less than a "D" Average	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Encouragement											
None	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
Once	0.0013	1,599	1	1,594	6	1,600	0	1,600	0	400	0
Twice	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
How Many Religious Services Attended in Past Year											
0 Times	0.0012	1,597	3	1,595	5	1,600	0	1,600	0	400	0
1–2 Times	0.0373	814	786	715	885	1,481	119	1,483	117	375	25
3–5 Times	0.1183	140	1,460	151	1,449	7	1,593	3	1,597	0	400
6–24 Times	0.0314	796	804	807	793	1,491	109	1,562	38	369	31
25–52 Times	0.4466	2	1,598	3	1,597	0	1,600	0	1,600	0	400
More than 52 Times	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Had Major Depressive Episode with Severe Role Impairment											
Yes	0.4657	1	1,599	4	1,596	0	1,600	0	1,600	0	400
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Substance Use Disorder											
No	0.8403	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Yes	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.19 Statistical Significance of Individual Covariates in N14/YOTMTHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Number of Visits/Overnight Stay for Specialty Mental Health Services											
1	0.1442	159	1,441	152	1,448	118	1,482	64	1,536	24	376
2	0.9667	0	1,600	0	1,600	0	1,600	0	1,600	0	400
3–6	0.5401	3	1,597	2	1,598	0	1,600	0	1,600	0	400
7–24	0.3831	6	1,594	5	1,595	0	1,600	0	1,600	0	400
25 or More	0.0053	1,497	103	1,467	133	1,508	92	1,588	12	381	19
None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Took Prescription Medication Prescribed for Mood During Past 12 Months											
Yes	0.0039	1,574	26	1,549	51	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.
N/A = not applicable. These rows correspond to reference levels of covariates.
Note: See Table 4.1 for study and model details.

Table F.20 Statistical Significance of Individual Covariates in N14/YORXHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Age											
12	0.4370	11	1,589	28	1,572	0	1,600	0	1,600	0	400
13	0.5206	3	1,597	1	1,599	0	1,600	0	1,600	0	400
14	0.2100	77	1,523	61	1,539	0	1,600	0	1,600	0	400
15	0.4603	1	1,599	2	1,598	0	1,600	0	1,600	0	400
16	0.8330	0	1,600	0	1,600	0	1,600	0	1,600	0	400
17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Gender											
Male	0.0286	915	685	919	681	1,538	62	1,588	12	394	6
Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity											
White	0.0457	367	1,233	458	1,142	1,085	515	1,144	456	307	93
Black/African American	0.6603	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Other	0.1148	130	1,470	190	1,410	0	1,600	0	1,600	0	400
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Income											
Less than $20,000	0.6506	3	1,597	3	1,597	0	1,600	0	1,600	0	400
$20,000–$49,999	0.0703	337	1,263	390	1,210	18	1,582	34	1,566	12	388
$50,000–$74,999	0.1686	104	1,496	81	1,519	0	1,600	0	1,600	0	400
$75,000–$99,999	0.2569	11	1,589	19	1,581	0	1,600	0	1,600	0	400
$100,000 or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Health Insurance											
Has Insurance	0.7226	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Does Not Have Insurance	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rural/Urban Segment											
Rural	0.4581	2	1,598	2	1,598	0	1,600	0	1,600	0	400
Urban	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Number of Delinquent Behaviors											
None	0.0005	1,600	0	1,600	0	1,600	0	1,600	0	400	0
One	0.0008	1,593	7	1,598	2	1,600	0	1,600	0	400	0
Two or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.20 Statistical Significance of Individual Covariates in N14/YORXHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Grade for Last Semester or Grading Period Completed											
An "A+," "A," or "A-Minus" Average	0.1686	52	1,548	56	1,544	42	1,558	9	1,591	17	383
A "B+," "B," or "B-Minus" Average	0.1638	46	1,554	59	1,541	46	1,554	12	1,588	16	384
A "C+," "C," or "C-Minus" Average	0.7507	0	1,600	0	1,600	0	1,600	0	1,600	0	400
A "D" or Less than a "D" Average	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Encouragement											
None	0.0006	1,598	2	1,598	2	1,600	0	1,600	0	400	0
Once	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
Twice	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
How Many Religious Services Attended in Past Year											
0 Times	0.0558	542	1,058	499	1,101	774	826	603	997	179	221
1–2 Times	0.2759	15	1,585	21	1,579	0	1,600	0	1,600	0	400
3–5 Times	0.7798	1	1,599	1	1,599	0	1,600	0	1,600	0	400
6–24 Times	0.0713	294	1,306	383	1,217	213	1,387	194	1,406	68	332
25–52 Times	0.1822	26	1,574	46	1,554	0	1,600	0	1,600	1	399
More than 52 Times	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Had Major Depressive Episode with Severe Role Impairment											
Yes	0.1246	136	1,464	171	1,429	5	1,595	1	1,599	2	398
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Substance Use Disorder											
No	0.3138	2	1,598	4	1,596	0	1,600	0	1,600	0	400
Yes	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.20 Statistical Significance of Individual Covariates in N14/YORXHLP by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	**N.S.**	**α** = 0.05	**N.S.**	**α** = 0.05	**N.S.**	**α** = 0.05	**N.S.**	**α** = 0.05	**N.S.**
Number of Visits/Overnight Stay for Specialty Mental Health Services											
1	0.7493	6	1,594	9	1,591	1	1,599	0	1,600	1	399
2	0.0254	977	623	957	643	953	647	1,122	478	250	150
3–6	0.2668	41	1,559	33	1,567	30	1,570	15	1,585	9	391
7–24	0.0021	1,590	10	1,581	19	1,592	8	1,600	0	396	4
25 or More	0.0422	861	739	754	846	787	813	850	750	208	192
None	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.

N/A = not applicable. These rows correspond to reference levels of covariates.

Note: See Table 4.1 for study and model details.

Table F.21 Statistical Significance of Individual Covariates in N19/ABODANL by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Had Past Year Major Depressive Episode											
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Support											
At Least 3	0.0022	1,580	20	1,540	60	1,578	22	1,599	1	394	6
4–5	0.2741	10	1,590	17	1,583	11	1,589	1	1,599	4	396
6–7	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age Group											
12–13	0.0399	496	1,104	526	1,074	603	997	1,237	363	151	249
14–15	0.4460	10	1,590	8	1,592	0	1,600	0	1,600	0	400
16–17	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Gender											
Male	0.0002	1,599	1	1,599	1	1,600	0	1,600	0	400	0
Female	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Race/Ethnicity											
White	0.7979	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Black/African American	0.7280	2	1,598	2	1,598	0	1,600	0	1,600	0	400
Other	0.8625	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Hispanic/Latino	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Family Income											
Less than \$20,000	0.4683	0	1,600	2	1,598	0	1,600	0	1,600	0	400
\$20,000–\$49,999	0.2937	11	1,589	29	1,571	0	1,600	0	1,600	0	400
\$50,000–\$74,999	0.8242	0	1,600	0	1,600	0	1,600	0	1,600	0	400
\$75,000 or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Rural/Urban Segment											
Rural	0.7493	0	1,600	0	1,600	0	1,600	0	1,600	0	400
Urban	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Alcohol Use Disorder											
Yes	0.0001	1,600	0	1,600	0	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Past Year Illicit Drug Use Disorder Excluding Pain Relievers											
Yes	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
No	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Table F.21 Statistical Significance of Individual Covariates in N19/ABODANL by MIVRA Method, Simulation versus Full Sample: 20 Percent Deletion Rate (continued)

Covariate	Full-Sample *P*-Value	LD		REWT		WSHD		MPLUS		CTBHD	
		α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.	α = 0.05	N.S.
Number of Delinquent Behaviors											
None	0.0000	1,600	0	1,600	0	1,600	0	1,600	0	400	0
One	0.0009	1,599	1	1,591	9	1,600	0	1,600	0	400	0
Two or More	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Grades for Last Semester											
An "A+," "A," or "A-Minus" Average	0.1548	54	1,546	60	1,540	39	1,561	8	1,592	13	387
A "B+," "B," or "B-Minus" Average	0.2038	26	1,574	14	1,586	15	1,585	2	1,598	6	394
A "C+," "C," or "C-Minus" Average	0.4989	0	1,600	0	1,600	0	1,600	0	1,600	0	400
A "D" or Less than a "D" Average	0.4551	0	1,600	0	1,600	0	1,600	0	1,600	0	400
School Does Not Give These Grades	0.6560	11	1,589	11	1,589	1	1,599	1	1,599	0	400
Not Enrolled in School	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
How Many Religious Services Attended in Past Year											
0 Times	0.0032	1,563	37	1,521	79	1,596	4	1,600	0	399	1
1–5 Times	0.0409	1,069	531	720	880	751	849	1,089	511	197	203
6–24 Times	0.0532	891	709	642	958	593	1,007	655	945	152	248
25–52 Times	0.0218	1,271	329	990	610	1,426	174	1,551	49	365	35
More than 52 Times	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

CTBHD = cyclical tree-based hot deck; LD = listwise deletion; MIVRA = missing item values in regression analyses; MPLUS = (pseudo-)maximum likelihood method using off-the-shelf Mplus® software; N.S. = not significant; REWT = listwise deletion with reweighting; WSHD = weighted sequential hot deck.
N/A = not applicable. These rows correspond to reference levels of covariates.
Note: See Table 4.1 for study and model details.

This page intentionally left blank