



2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book

Section 10: Editing and Imputation Report

2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 10: Editing and Imputation Report

This publication was prepared for the Substance Abuse and Mental Health Services Administration (SAMHSA) under contract number 75S20322C00001 with SAMHSA, U.S. Department of Health and Human Services (HHS). Marlon Daniel and Carlos Graham served as contracting officer representatives.

Recommended Citation

Center for Behavioral Health Statistics and Quality. (2025). *2023 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 10: Editing and imputation report*. <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/methodology>

Originating Office

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, 5600 Fishers Lane, Rockville, MD 20857. Released 2025.

Electronic Access

Products may be downloaded at <https://library.samhsa.gov/>.

Disclaimer

Nothing in this document constitutes a direct or indirect endorsement by SAMHSA or HHS of any nonfederal entity's products, services, or policies.

Public Domain Notice

This publication is in the public domain and may be reproduced or copied without permission from SAMHSA. Citation of the source is appreciated. However, this publication may not be reproduced or distributed for a fee without the specific, written authorization of the Office of Communications, SAMHSA, HHS.

U.S. Department of Health and Human Services
Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Office of Population Surveys

Released 2025

Contents

1.	<u>Introduction</u>	<u>1</u>
1.1	<u>Organization of the NSDUH Questionnaire and Overview of Content</u>	<u>1</u>
1.2	<u>Overall Data Quality Issues with CAI.....</u>	<u>7</u>
1.3	<u>Organization of This Report.....</u>	<u>8</u>
1.4	<u>Changes from the 2022 Survey to the 2023 Survey</u>	<u>10</u>
1.4.1	<u>Changes to Accommodate Modifications to the Questionnaire.....</u>	<u>10</u>
1.4.2	<u>Changes Involving Minor Corrections and Enhancements.....</u>	<u>11</u>
2.	<u>Procedures and General Principles for Editing the NSDUH Data.....</u>	<u>13</u>
2.1	<u>Introduction</u>	<u>13</u>
2.2	<u>Initial Processing of NSDUH Interview Data.....</u>	<u>13</u>
2.2.1	<u>Assignment of Standard NSDUH Missing Data Codes.....</u>	<u>14</u>
2.2.2	<u>Remapping of Responses to “Enter All That Apply” Questions</u>	<u>14</u>
2.2.3	<u>Identification of Usable Interviews</u>	<u>16</u>
2.2.4	<u>Creation of Cumulative Quarterly Unedited Data Files.....</u>	<u>18</u>
2.3	<u>Coding of “OTHER, Specify” Data and Investigation of Response Patterns.....</u>	<u>18</u>
2.3.1	<u>Coding of “OTHER, Specify” Data.....</u>	<u>18</u>
2.3.2	<u>Investigation of Response Patterns in Usable Interviews</u>	<u>27</u>
2.4	<u>General Issues and Principles for Editing NSDUH Data</u>	<u>31</u>
2.4.1	<u>Editing across Sections.....</u>	<u>31</u>
2.4.2	<u>Assignment of Relevant “Not Applicable” Codes</u>	<u>32</u>
2.4.3	<u>Additional Assignment of NSDUH Missing Data Codes as Part of Editing within a Section</u>	<u>35</u>
2.4.4	<u>Assignment of Codes to Edited “Enter All That Apply” Variables</u>	<u>35</u>
2.4.5	<u>“Back-Editing” Based on “OTHER, Specify” Data</u>	<u>37</u>
2.4.6	<u>“Lockout” Following Completion of Relevant Sections</u>	<u>39</u>
2.4.7	<u>“Flag and Impute” Principles for Substance Use Data Other than Prescription Drugs.....</u>	<u>40</u>
3.	<u>Imputation Procedures Used in NSDUH</u>	<u>42</u>
3.1	<u>Introduction</u>	<u>42</u>
3.2	<u>Development of the Predictive Mean Neighborhood Methodology</u>	<u>46</u>
3.3	<u>Implementation of the Predictive Mean Neighborhood Methodology</u>	<u>47</u>

3.3.1	Step 1: Response Propensity Adjustment.....	47
3.3.2	Step 2: Prediction Modeling	48
3.3.3	Step 3: Hot-Deck Imputation	52
3.4	Predictive Mean Neighborhood “Types”	57
3.4.1	Type 1: Single Response Propensity/Single Prediction	58
3.4.2	Type 2: Multiple Response Propensity/Multiple Prediction	59
3.4.3	Type 3: Single Response Propensity/Multiple Prediction.....	61
3.5	Development of the Modified Predictive Mean Neighborhood Methodology	62
3.6	Procedural Steps for the Modified Predictive Mean Neighborhood Methodology	64
3.6.1	Step 1: Response Propensity Adjustment.....	64
3.6.2	Step 2: Prediction Modeling	65
3.6.3	Step 3: Imputation.....	68
3.7	Implementation of the Modified Predictive Mean Neighborhood Methodology	70
3.7.1	Sequencing of Variables for modPMN	73
3.7.2	Application of modPMN for a Single Variable.....	74
3.7.3	Cycling in modPMN	75
3.8	Special Auxiliary Variables: Age Group and State Rank.....	77
3.8.1	Age Groups	77
3.8.2	State Rank	78
3.9	Modeling Techniques to Improve Imputation Efficiency.....	78
4.	Editing and Imputation for the NSDUH Front-End Demographics Variables.....	80
4.1	Introduction	80
4.2	Editing Selected Front-End Demographics Variables.....	81
4.2.1	Creating the Edited Interview Date Variable (INTDATE)	81
4.2.2	Creating the Edited Age Variable (AGE).....	82
4.2.3	Creating the Edited Current State Residency Variable (STATELOC)	85
4.2.4	Creating the Edited Variables for Sex Assigned at Birth (IRSEX) and Gender Identity.....	86
4.2.5	Creating the Edited Birth Date Variable (BRTHDATE).....	87
4.2.6	Creating the Edited Marital Status Variables (MARITAL, EDMARIT, and NOMARR)	87

4.2.7	Creating the Edited Race and Hispanic/Latino (Origin and Group) Variables	88
4.2.8	Creating the Edited Military Service History and Military Veteran Status Variables	95
4.2.9	Creating the Edited Highest Grade Completed Variables (EDUHIGHST and EDEDUC)	97
4.2.10	Creating the Perceived Health Status Variable (HEALTH)	99
4.2.11	Creating the Interviewer Debriefing Variables	99
4.3	Imputation for Selected Front-End Demographics Variables	99
4.3.1	Marital Status (Imputation Set 1)	99
4.3.2	Race Variables (Imputation Set 2)	100
4.3.3	Hispanic/Latino Origin Variable (Imputation Set 3)	103
4.3.4	Hispanic/Latino Group Variable (Imputation Set 4)	103
4.3.5	Education Level Variable (Imputation Set 5)	105

5. Editing and Imputation for the NSDUH Back-End Demographics Variables ... 108

5.1	Introduction	108
5.2	Editing the Back-End Demographics Variables	108
5.2.1	Editing for Patterned Responses	108
5.2.2	Moves in the Past 12 Months	108
5.2.3	Immigrant Status	109
5.2.4	Sexual Attraction and Identity	110
5.2.5	English Language Proficiency and Disability	112
5.2.6	Current Education	112
5.2.7	Military Families	121
5.2.8	Employment and Workplace	122
5.3	Imputation for Back-End Demographics Variables	125
5.3.1	Employment Status (Imputation Set 1)	125
5.3.2	Immigrant Status, Born-in-U.S. Indicator (Imputation Set 2)	127
5.3.3	Immigrant Status, Age of Entry (Imputation Set 3)	127

6. Editing and Imputation for the NSDUH Substance Use Variables 129

6.1	Introduction	129
6.1.1	Edited (but Not Imputed) Substance Use Variables	129
6.1.2	"Parent" and "Child" Substance Use Measures	130
6.2	Editing Procedures for the Substance Use Variables Other Than Prescription Drugs	131

6.2.1	Edits of Lifetime Use Variables	133
6.2.2	Edits of Recency-of-Use Variables for Substances Other Than Prescription Drugs.....	138
6.2.3	Edits Involving Users of Only Child Drugs for Hallucinogens.....	147
6.2.4	Edits of Frequency-of-Use Variables	149
6.2.5	Initiation of Use (Age at First Use, Month of First Use, and Year of First Use)	154
6.2.6	Editing of Lifetime Daily Cigarette Use Status	156
6.2.7	Edits for Tobacco Variables That Do Not Undergo Imputation.....	157
6.2.8	Edits for Additional Marijuana Variables.....	158
6.3	Editing Procedures for the Prescription Drug Variables	165
6.3.1	Mentions of Over-the-Counter Drugs.....	167
6.3.2	Any Past Year or Lifetime Use of Specific Prescription Drugs.....	168
6.3.3	Most Recent Use of Prescription Drugs for Any Reason.....	172
6.3.4	Misuse of Specific Prescription Drugs in the Past 12 Months	173
6.3.5	Most Recent Misuse of Prescription Drugs.....	177
6.3.6	Initiation of Misuse of Prescription Drugs.....	178
6.3.7	Misuse of Prescription Drugs in the Past 30 Days	183
6.3.8	Ways of Misuse in the Past 12 Months.....	186
6.3.9	Injection of Prescription Stimulants in the Past 12 Months or Past 30 Days	188
6.3.10	Identification of the Last Prescription Drug Misused in the Past 12 Months	189
6.3.11	Reasons for the Last Misuse of Prescription Drugs in the Past 12 Months	190
6.3.12	Sources of Prescription Drugs for the Last Misuse in the Past 12 Months	194
6.3.13	Edit Rules for Prescription Drugs.....	195
6.4	Additional Edits Applied during Imputation Processing	197
6.4.1	“Other” Hallucinogens.....	198
6.4.2	Respondents Imputed to Lifetime Use for Child Drug Variables.....	198
6.4.3	Age-at-First-Use Variables	199
6.4.4	Benzodiazepines Variables.....	200
6.4.5	Fentanyl Products Variables.....	200
6.5	Imputation of Substance Use Variables	201
6.5.1	Variables Imputed Using PMN Methodology.....	201
6.5.2	Variables Imputed Using the modPMN Methodology.....	228
6.5.3	Recodes for Additional Analyses.....	248

7.	<u>Editing and Imputation for Special Drugs and Substance Use Disorder Variables.....</u>	<u>250</u>
7.1	<u>Introduction</u>	<u>250</u>
7.2	<u>Special Drugs</u>	<u>250</u>
7.2.1	<u>GHB and OTC Cold Medicines</u>	<u>250</u>
7.2.2	<u>General Edits of Cocaine, Heroin, and Methamphetamine Variables for Nonusers</u>	<u>252</u>
7.2.3	<u>Edits for Heroin Users Who Did Not Report Smoking, Sniffing, or Injection of Heroin</u>	<u>253</u>
7.2.4	<u>Additional Edits for Smoking or Sniffing of Heroin</u>	<u>253</u>
7.2.5	<u>Injection Drug Use</u>	<u>254</u>
7.2.6	<u>Imputation of Variables for GHB and OTC Cold Medicines.....</u>	<u>259</u>
7.3	<u>Substance Use Disorder</u>	<u>262</u>
7.3.1	<u>Editing of Substance Use Disorder and Cigarette Dependence Variables</u>	<u>262</u>
7.3.2	<u>Imputation of Cigarette Dependence and Substance Use Disorder Variables.....</u>	<u>273</u>
8.	<u>Editing and Imputation in Additional Sections for Substance Use.....</u>	<u>280</u>
8.1	<u>Introduction</u>	<u>280</u>
8.2	<u>Blunts</u>	<u>280</u>
8.3	<u>Prior Substance Use.....</u>	<u>283</u>
8.3.1	<u>Retrospective Use</u>	<u>284</u>
8.3.2	<u>Last Use of Drugs</u>	<u>284</u>
8.4	<u>Alcohol and Drug Use Treatment</u>	<u>287</u>
8.4.1	<u>Type and Location of Alcohol and Drug Treatment.....</u>	<u>287</u>
8.4.2	<u>Treatment for Specific Substances</u>	<u>289</u>
8.4.3	<u>Other Alcohol or Drug Use Services.....</u>	<u>293</u>
8.4.4	<u>Perceived Unmet Need for Alcohol and Drug Use Treatment.....</u>	<u>295</u>
8.4.5	<u>Reasons for Not Receiving Alcohol or Drug Use Treatment</u>	<u>296</u>
8.4.6	<u>Imputation</u>	<u>296</u>
8.5	<u>Consumption of Alcohol</u>	<u>298</u>
8.5.1	<u>General Skip Logic Edits for Alcohol-Related Variables.....</u>	<u>299</u>
8.5.2	<u>Number of Drinks for the Last Alcohol Use in the Past 30 Days</u>	<u>299</u>
8.5.3	<u>Source and Location of Last Underage Alcohol Use in the Past 30 Days.....</u>	<u>299</u>
8.5.4	<u>Use of Illicit Drugs with Alcohol in the Past 30 Days</u>	<u>300</u>

8.5.5	Binge Alcohol History	302
8.6	Emerging Issues.....	304
8.6.1	Perceived Recovery from Substance Use or Mental Health Problems.....	304
8.6.2	Kratom.....	305
8.6.3	Vaping of Only Flavoring	305
8.6.4	Synthetic Marijuana and Synthetic Stimulants	305
8.6.5	Illegally Made Fentanyl.....	305
8.6.6	Imputation	306
8.7	Market Information for Marijuana	309
9.	Editing of Risk and Protective Factors for Substance Use Variables	312
9.1	Introduction	312
9.2	Risk/Availability	312
9.3	Special Topics	313
9.3.1	Editing of Variables for Criminal Justice Involvement.....	313
9.3.2	Editing of Variables for Driving under the Influence.....	314
9.3.3	Editing of Variables for State Marijuana Laws	318
9.4	Social Environment.....	318
9.5	Youth Experiences	319
9.5.1	Suicidal Thoughts and Behavior among Youths	320
9.6	Parenting Experiences.....	320
9.6.1	Editing of the Checkpoint Variables	322
9.6.2	Editing of Variables in the Parenting Experiences Section	326
10.	Editing and Imputation of Sections for Physical and Mental Health	330
10.1	Introduction	330
10.2	Health.....	330
10.2.1	Pregnancy Status	331
10.2.2	Height and Weight.....	331
10.2.3	Utilization of Medical Services	332
10.2.4	Screening for Substance Use or Talks about Substance Use with a Healthcare Professional	333
10.2.5	Specific Health Conditions	337
10.3	Mental Health Services Utilization	339

10.3.1	Type and Location of Mental Health Treatment or Other Services	340
10.3.2	Perceived Unmet Need for Mental Health Treatment	341
10.3.3	Reasons for Not Receiving Mental Health Treatment	342
10.4	Adult Mental Health	343
10.5	Adult and Adolescent Depression	343
10.6	COVID-19	348
10.6.1	Editing for Skip Logic on the Perceived Effect of the COVID-19 Pandemic on Alcohol Use.....	349
10.6.2	Skip Logic for Effect of COVID-19 Pandemic on Other Drug Use.....	349
10.7	Imputation	350
10.7.1	Mental Health Services Utilization.....	350
10.7.2	Adult Mental Health and Adult Depression Variables.....	352
10.7.3	Recodes for Additional Analyses.....	359
11.	Editing and Imputation for the NSDUH Roster Variables	360
11.1	Introduction	360
11.2	Editing the Respondent-Level Detailed Roster Variables	363
11.2.1	Roster Consistency Checks	364
11.2.2	Roster-Level Dataset	365
11.2.3	Roster Edits Involving the Self	365
11.2.4	Roster Edits for Other Household Members	367
11.2.5	Final Edited Respondent-Level Detailed Roster Variables	369
11.3	Editing and Imputation Procedures for Roster-Derived Household Composition Variables.....	369
11.3.1	Creation of Edited Roster-Derived Household Composition Variables	369
11.3.2	Imputation for Roster-Derived Household Composition Variables ...	371
11.4	Editing the Proxy Variables.....	372
11.4.1	Edited Indicator of Potential Proxies in Household (EDFAM18).....	375
11.4.2	Editing the Proxy Variables when EDFAM18 = 1.....	375
11.4.3	Editing the Proxy Variables when EDFAM18 = 0.....	376
11.4.4	Additional Editing for PRXRELAT	376
11.4.5	Additional Editing for QP03A.....	376
12.	Editing and Imputation for the NSDUH Income Variables	377
12.1	Introduction	377

12.2	Editing for Binary and Finer Category Income Variables	378
12.2.1	Binary Income Variables.....	378
12.2.2	Finer Category Income Variables.....	379
12.2.3	Additional Edits to Income Variables for Inconsistencies or Logical Inferences.....	380
12.3	Editing for Cell Phone Variables	381
12.4	Imputation for the Income Variables.....	381
12.4.1	Imputation for Binary Income Variables (Imputation Set 1)	382
12.4.2	Imputation for Finer Category Income Variables (Imputation Set 2).....	386
13.	Editing and Imputation for the NSDUH Health Insurance Variables.....	390
13.1	Introduction	390
13.2	Editing the Health Insurance Variables.....	390
13.2.1	Editing Because of Incorrect State Information	391
13.2.2	Determination of Any Health Insurance Coverage	393
13.2.3	Editing of Health Insurance Follow-Up Variables Based on Current Coverage.....	394
13.2.4	Miscellaneous Issues for Editing the Health Insurance Variables	395
13.2.5	Editing Types of Health Insurance and Any Other Health Insurance Variables, Imputation Sets 1 and 2.....	396
13.3	Imputation for the Health Insurance Variables.....	399
13.3.1	Types of Health Insurance Variables, Imputation Set 1.....	399
13.3.2	Any Other Health Insurance Variable, Imputation Set 2.....	401
13.3.3	Health Insurance Recodes for Additional Analyses.....	402
14.	Editing and Imputation for the NSDUH Pair Variables	403
14.1	Introduction	403
14.1.1	Pair Relationship Variables.....	403
14.1.2	Multiplicity Count Variables	404
14.1.3	Household-Level Person Count Variables	406
14.2	Stage One Editing: Pair Relationships.....	407
14.2.1	Editing the Household Roster of Each Pair Member	407
14.2.2	Creating the Pair Relationship Variable (PAIRREL).....	407
14.3	Stage Two Editing: Multiplicity Counts.....	413
14.3.1	Determining the Multiplicity Count for Each Pair Member.....	414
14.3.2	Determining the Final Multiplicity Count.....	415

<u>14.4</u>	<u>Stage Three Editing: Household-Level Person Counts.....</u>	<u>416</u>
<u>14.4.1</u>	<u>Determining the Household-Level Person Count for Each Respondent</u>	<u>417</u>
<u>14.4.2</u>	<u>Determining the Final Household-Level Person Count.....</u>	<u>422</u>
<u>14.5</u>	<u>Imputation for the Pair Variables</u>	<u>423</u>
<u>14.5.1</u>	<u>Creation of Covariates for Imputing Pair-Level Variables.....</u>	<u>423</u>
<u>14.5.2</u>	<u>Stage One Imputation: Pair Relationships.....</u>	<u>426</u>
<u>14.5.3</u>	<u>Stage Two Imputation: Multiplicity Counts.....</u>	<u>429</u>
<u>14.5.4</u>	<u>Stage Three Imputation: Household-Level Person Counts.....</u>	<u>431</u>
References		435
Acknowledgments		439

Figures

3.1	Donor Selection Algorithm	53
3.2	PMN Type 1: Single Response Propensity/Single Prediction	58
3.3	PMN Type 2: Multiple Response Propensity/Multiple Prediction.....	60
3.4	PMN Type 3: Single Response Propensity/Multiple Prediction	62
3.5	Stochastic Imputation for a Three-Level Categorical Variable	69
3.6	Twelve-Step Block of modPMN Programs.....	75
3.7	Arranging the 12-Step Blocks to Complete Imputation for an Imputation Set	76
5.1	Follow-Up Question Logic for Current Education.....	113
5.2	Routing Logic for Respondents in the Employment Status Section	123
6.1	Routing Logic for Respondents in the Substance Use Sections Other Than for Prescription Drugs.....	133
6.2	Routing Logic for Respondents in the Prescription Psychotherapeutic Drug Sections	166
13.1	Routing Logic in the Health Insurance Section of the Questionnaire	392

Tables

1.1	Content of Demographic Sections in the 2023 NSDUH Interview	3
1.2	Content of Initial Substance Use Sections in the 2023 NSDUH Interview	3
1.3	Content of Sections on Remaining Topics in the 2023 NSDUH Interview	5
1.4	Content of Chapters	8
1.5	Content of Appendices.....	9
1.6	Comparison of Imputed Marijuana Vaping Estimates from the 2022 and 2023 Detailed Tables	12
2.1	Questionnaire Sections Screened in Data Diagnostic Review	27
2.2	Response Patterns for Pain Reliever Screener Question PR03.....	37
3.1	Weighted Percentages of Cases Imputed by Variable Group.....	43
3.2	Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood	50
3.3	Values of Delta for Various Predicted Probabilities	54
3.4	Examples of Variables Imputed Using Each of the Four Combinations of Univariate/Multivariate Matching and Assignment.....	56
3.5	PMN Types Applied to Each Variable Group and Imputation Set	58
3.6	Regression Models Used for Each Variable Imputed Using modPMN	65
3.7	Imputation Sets Processed Using modPMN	71
3.8	Example of Sequencing for Imputation Sets 10 and 21	74

3.9	<u>Independent Variables Used in Response Propensity and Prediction Steps of modPMN.....</u>	<u>75</u>
4.1	<u>Rules to Create the Intermediate Value for Age.....</u>	<u>84</u>
4.2	<u>Edited Variable Name for Each Race Group.....</u>	<u>91</u>
4.3	<u>Hispanic/Latino Status of Item Nonrespondents for Race</u>	<u>102</u>
4.4	<u>Edited Race Variables and Their Imputation-Revised Counterparts</u>	<u>102</u>
4.5	<u>Mapping of EDHOGRP Levels to EDHOGRP2 Levels.....</u>	<u>104</u>
4.6	<u>Mapping of EDEDUC Levels to RESPEDUC Levels.....</u>	<u>106</u>
5.1	<u>Names for Current Education Variables.....</u>	<u>114</u>
5.2	<u>Mapping of Current Age with AccepRange of Grades</u>	<u>117</u>
5.3	<u>Names for Military Family Variables</u>	<u>121</u>
5.4	<u>Categories of WRKSTATWK.....</u>	<u>124</u>
6.1	<u>Substance Use Measures with Parent/Child Relationships.....</u>	<u>131</u>
6.2	<u>How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables.....</u>	<u>143</u>
6.3	<u>How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables</u>	<u>145</u>
6.4	<u>States Where a Law or Initiative Allowing Marijuana for Medical Use was Passed as of December 31, 2023.....</u>	<u>160</u>
6.5	<u>“OTHER, Specify” Codes for Reasons for Misuse That Corresponded to Available Response Options</u>	<u>191</u>
6.6	<u>Inconsistent Patterns in the Prescription Drug Modules and Their Resolution</u>	<u>196</u>
6.7	<u>Substances and Substance Use Measures, PMN Imputation Sets</u>	<u>202</u>
6.8	<u>General Incomplete Recency Categories for Cigarettes</u>	<u>207</u>
6.9	<u>Elements of Full Predictive Mean Vector: Cigarettes.....</u>	<u>209</u>
6.10	<u>Elements of Full Predictive Mean Vector.....</u>	<u>223</u>
6.11	<u>Substances and Substance Use Measures, modPMN Imputation Sets.....</u>	<u>229</u>
6.12	<u>Sequencing of Substances within Imputation Sets, 2023.....</u>	<u>230</u>
6.13	<u>Summary of Response Propensity Models for Categorical modPMN Variables</u>	<u>232</u>
6.14	<u>Outcomes of Prediction Modeling</u>	<u>233</u>
6.15	<u>Final Imputation-Revised Categorical modPMN Variables</u>	<u>235</u>
6.16	<u>Summary of Response Propensity Models for Continuous modPMN Variables</u>	<u>241</u>
6.17	<u>Final Imputation-Revised Continuous modPMN Variables</u>	<u>243</u>
6.18	<u>Prevalence Recodes.....</u>	<u>248</u>
6.19	<u>Prevalence Recodes Incorporating More Than One Recency Variable.....</u>	<u>249</u>
7.1	<u>Substances and Substance Use Measures</u>	<u>260</u>
7.2	<u>Sequencing of Substances within Imputation Sets, 2023.....</u>	<u>260</u>
7.3	<u>Summary of Response Propensity Models</u>	<u>260</u>

7.4	Outcomes of Prediction Modeling	261
7.5	Prevalence Recodes.....	261
7.6	Mapping of Cigarette Dependence Questions to Edited Variables.....	264
7.7	Quick Reference Guide to Creation of Edited DSM-5 Disorder Indicators	267
7.8	DSM-5 Substance Use Disorder Diagnosis Criteria.....	270
7.9	DSM-5 Substance Use Disorder Diagnosis Determined by Number of Positive Criteria	272
7.10	Segregation of Substances into Imputation Sets.....	275
7.11	Sequencing of DSM-5 Disorder Indicator and Disorder Severity within Imputation Sets, 2023	276
7.12	DSM-5 Recoded Variables	278
8.1	Substance Use Treatment Variables in Imputation Set 1	297
8.2	Substance Use Treatment Variables in Imputation Set 2	297
8.3	Substances and Substance Use Measures	306
8.4	Sequencing of Substances within Imputation Sets, 2023.....	306
8.5	Summary of Response Propensity Models	307
8.6	Outcomes of Prediction Modeling	308
8.7	Prevalence Recodes.....	309
9.1	Mapping of Edited Variables for Suicidal Thoughts and Behavior among Youths	320
9.2	Parenting Experiences Checkpoint Questions by Interview Mode	322
10.1	Depression Symptom and Score Variables.....	346
10.2	Mental Health Services Utilization Variables in Imputation Sets 1 and 2	351
10.3	Imputation Sets	353
10.4	Sequencing of Outcomes within Multiple-Variable Imputation Sets	354
10.5	Summary of Response Propensity Models	354
10.6	Edited Kessler-6 Variables.....	356
10.7	Edited WHODAS Variables	357
10.8	Imputation-Revised Variables.....	359
10.9	Recoded Variables.....	359
11.1	Roster Grid Example Where Number of Individuals in Household (QD54) Equals 4	360
11.2	Detailed Roster Relationship Codes (2016-Present) and Equivalent Former Roster Relationship Codes (2002-2015).....	361
11.3	Roster-Derived Household Composition Variables	362
11.4	Proxy Variables	363
11.5	Relationship Codes for Each Possible Person of Reference	367
11.6	Changes to Roster Age or Gender Based on the Other Pair Member Reported.....	369
11.7	Levels of PRXRELAT	373
11.8	Levels of PRXRELATD	374

<u>12.1</u>	<u>Comparison between Personal Level and Family Level of Income Questions.....</u>	<u>377</u>
<u>12.2</u>	<u>Imputation Order for Binary Income Variables (Imputation Set 1)</u>	<u>382</u>
<u>12.3</u>	<u>Cross-Classification of FAMPMT and FAMSVCS; Relevant Combinations for Binary Income Predictive Mean Vector when WELMOS Is Missing</u>	<u>385</u>
<u>13.1</u>	<u>Edits Pertaining to the Health Insurance Section</u>	<u>395</u>
<u>13.2</u>	<u>Mapping of Raw Health Insurance Variables to Edited Variables and Base Variables for Imputation</u>	<u>397</u>
<u>14.1</u>	<u>Edited and Imputation-Revised Multiplicity Count Variables.....</u>	<u>405</u>
<u>14.2</u>	<u>Edited and Imputation-Revised Household-Level Person Count Variables</u>	<u>407</u>
<u>14.3</u>	<u>Levels of the Variable PAIRREL</u>	<u>408</u>
<u>14.4</u>	<u>Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH</u>	<u>412</u>
<u>14.5</u>	<u>Frequencies of the Levels of the Variable RELMATCH: 2023</u>	<u>413</u>
<u>14.6</u>	<u>Multiplicity Counts for Each Pair Member</u>	<u>415</u>
<u>14.7</u>	<u>Amount of Imputation Required for Multiplicities in Various Pair Domains: 2023.....</u>	<u>416</u>
<u>14.8</u>	<u>Obtaining the Bounds for the Parent-Child, Child (12-14) Domain</u>	<u>418</u>
<u>14.9</u>	<u>Obtaining the Bounds for the Sibling-Sibling (12-14/15-17) Domain</u>	<u>420</u>
<u>14.10</u>	<u>Obtaining the Bounds for the Spouse-Spouse Domain</u>	<u>421</u>
<u>14.11</u>	<u>Obtaining the Bounds for the Spouse-Spouse with Children Younger Than 18 Domain</u>	<u>422</u>
<u>14.12</u>	<u>Conditions Used to Reconcile HHSIZE across Pairs.....</u>	<u>424</u>
<u>14.13</u>	<u>Conditions Used to Reconcile the Age Counts across Pairs</u>	<u>425</u>
<u>14.14</u>	<u>Age Group Pairs with Associated Possible Pair Relationships.....</u>	<u>426</u>
<u>14.15</u>	<u>Modeled Pair Relationships within Age Group Pairs</u>	<u>429</u>
<u>14.16</u>	<u>Modeled Multiplicities within Domains</u>	<u>430</u>

1. Introduction

Conducted annually, the National Survey on Drug Use and Health (NSDUH) is the primary source of information on alcohol use, tobacco use, illicit drug use (including misuse of prescription drugs), substance use disorders, and mental health issues for the U.S. civilian, noninstitutionalized population aged 12 years or older. This population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. The target population excluded people with no fixed household address (e.g., homeless or transient people not in shelters), residents of institutional group quarters (e.g., jails and nursing homes), children younger than 12 years of age, and active military personnel. People also were excluded during data collection if they could not complete the survey in either English or Spanish or they were not physically or mentally capable of completing the interview. The 2023 NSDUH (conducted by RTI International¹) continued to utilize a stratified multistage area probability sample design that enables the Center for Behavioral Health Statistics and Quality within the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide representative estimates of the nation as a whole and for each of the 50 states and the District of Columbia (see Appendix A for more details). Both direct and model-based state and substate estimates are produced on a variety of measures based on a combination of multiple years of data.

The 2023 NSDUH continued to use multimode data collection (in-person and web-based data collection). Methods for multimode data collection in the 2023 NSDUH are provided in the *2023 National Survey on Drug Use and Health (NSDUH): Methodological Summary and Definitions* (Center for Behavioral Health Statistics and Quality, 2024a).

This report focuses on the editing and statistical imputation procedures that were applied to respondent data for the 2023 NSDUH. Reports that document editing and imputation procedures in prior years are available on the [NSDUH Data Collection](#) web page. Logical editing uses data from elsewhere within the same respondent's record to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables. Imputation is defined as the replacement of missing values with valid, nonmissing values. To preserve the natural variability in the data, statistical imputation usually involves some randomness.

1.1 Organization of the NSDUH Questionnaire and Overview of Content

Since 1999, NSDUH data have been collected using computer-assisted interviewing (CAI).² The in-person interview is organized by sections based on content, routing logic, and the mode of administration, either self-administered via audio computer-assisted self-interviewing (ACASI) or interviewer administered via computer-assisted personal interviewing (CAPI). Web-based data collection required some modifications to data collection procedures to accommodate the web

¹ RTI International is a trade name of Research Triangle Institute. RTI and the RTI logo are U.S. registered trademarks of Research Triangle Institute.

² CAI specifications for the 2023 NSDUH are available on the [NSDUH Data Collection](#) web page.

environment. The web interview was also organized by section based on content and routing logic, but all questions were self-administered. However, the content of questions and question order were generally the same for in-person and web-based data collection. For in-person and web interviews, the logic for determining which questions respondents should be asked was controlled by the computer program based on the respondents' answers to previous questions.

In-person data collection was completed using laptop computers. Respondents completing the survey via the web could do so from any device with Internet access (e.g., smartphone, tablet, computer).

The first section of the 2023 interview (core demographics) consisted of questions about certain demographic characteristics, including birth date (which is used to determine a respondent's age) and education level (highest grade completed) ([Table 1.1](#)). For in-person interviews, CAPI was used for these questions, a process in which interviewers read the questions and respondents gave their answers aloud to the interviewers, who then entered the responses into the computer.

Following completion of these demographic questions, interviewers oriented in-person respondents about the use of specific features of the laptop computer and use of headphones for listening to questions. Respondents completing the in-person interview then completed a brief tutorial on answering self-administered questions, which were administered through ACASI. ACASI is used for most questions in the in-person interview and was particularly useful for collecting data about sensitive topics such as (but not limited to) substance use, problems associated with substance use, risk and protective factors related to substance use, and mental health issues.

In the ACASI portion of the interview, in-person respondents could read the questions on a computer screen and also were encouraged to listen to an audio recording of the questions on headphones. In-person respondents then entered their answers directly into a laptop computer. These features of ACASI prevented interviewers or others in the household from knowing what questions the respondents were asked and how they were answering. The availability of audio recordings of the questions was especially useful for respondents with limited reading ability.

Because web interviews were entirely self-administered, web-based data collection did not use ACASI, and audio recordings of questions were not available for web respondents. To complete the survey via the web, therefore, web respondents needed to be able to read English or Spanish and to see the questions on whatever device they were using. Because the web questionnaire did not include audio, pronunciations were spelled out visually on several screens, particularly for hallucinogens, inhalants, and prescription drug introduction screens, to help youths and respondents with a lower reading level understand the questions accurately.

Beginning in 2023, questions about demographic characteristics such as sex assigned at birth, gender identity, Hispanic/Latino origin, racial group, overall health, and marital status were moved from the core demographics section to a new front-end demographics section where questions were self-administered for all respondents ([Table 1.1](#)). Prior to 2023, questions regarding these demographic characteristics were asked in the first demographic section using

CAPI for in-person respondents. Also prior to 2023, there was a single question asking respondents whether they were male or female. Beginning in 2023 respondents were asked two separate questions about (1) sex assigned at birth and (2) gender identity.

Table 1.1 Content of Demographic Sections in the 2023 NSDUH Interview

Section	Content
Core Demographics ¹	▪ Birth date, state of residence, military status, and education (highest grade completed)
Front-End Demographics	▪ Hispanic/Latino origin, racial group, marital status, sex assigned at birth, gender identity, and overall health
Back-End Demographics ²	▪ Moves in the past 12 months and state residency 1 year ago (if moved in the past 12 months) ▪ Immigrant status and length of time living in the United States ▪ Sexual attraction and sexual identity ▪ English language proficiency and disability
Education ²	▪ Current school enrollment ▪ Military families
Employment ²	▪ Current employment status ▪ Employment in the past 12 months (if not currently employed) ▪ Absences from work in the past 30 days ▪ Substance use education and testing policies in the workplace

¹ Questions are interviewer administered for in-person respondents.

² Questions are asked toward the end of the interview, following sections regarding substance use and mental health issues.

Following the demographic questions from the core demographic and front-end demographic sections, respondents were asked questions about their substance use; the substance use questions were the first set of self-administered questions for in-person respondents. [Table 1.2](#) shows questionnaire sections on the use of nicotine (including tobacco products), alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, methamphetamine, and prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives).

Table 1.2 Content of Initial Substance Use Sections in the 2023 NSDUH Interview

Section	Content
Nicotine	▪ Lifetime use or nonuse of cigarettes, nicotine vaping, smokeless tobacco, cigars, and pipe tobacco ▪ Additional questions for lifetime users
Alcohol	▪ Lifetime use or nonuse ▪ Additional questions for lifetime users, including binge drinking for past month users
Marijuana	▪ Lifetime use or nonuse of CBD or hemp products and marijuana or any cannabis product ▪ Additional questions for lifetime users, including modes of marijuana use for past year and past month users and medically approved use of marijuana
Cocaine and Crack	▪ Lifetime use or nonuse of any cocaine and crack cocaine (if lifetime cocaine user) ▪ Additional questions for lifetime users

Table 1.2 Content of Initial Substance Use Sections in the 2023 NSDUH Interview (continued)

Section	Content
Heroin	▪ Lifetime use or nonuse ▪ Additional questions for lifetime users
Hallucinogens	▪ Lifetime use or nonuse of 10 types of hallucinogens, including any other hallucinogen besides the ones that had been listed ▪ Additional questions if lifetime use was reported for any of the 10 types of hallucinogens
Inhalants	▪ Lifetime use or nonuse of 13 specific types of inhalants for kicks or to get high, including any other inhalant besides the ones that had been listed ▪ Additional questions if lifetime use was reported for any of the 13 types of inhalants
Methamphetamine	▪ Lifetime use or nonuse ▪ Additional questions for lifetime users
Pain Relievers Screener	▪ Any use in the past 12 months of 36 specific pain relievers and any other pain reliever ▪ Any lifetime use of prescription pain relievers, if past year use was not reported
Tranquilizers Screener	▪ Any use in the past 12 months of 12 specific tranquilizers and any other tranquilizer ▪ Any lifetime use of prescription tranquilizers, if past year use was not reported
Stimulants Screener	▪ Any use in the past 12 months of 26 specific stimulants and any other stimulant ▪ Any lifetime use of prescription stimulants, if past year use was not reported
Sedatives Screener	▪ Any use in the past 12 months of 14 specific sedatives and any other sedative ▪ Any lifetime use of prescription sedatives, if past year use was not reported
Pain Relievers Main Section	▪ Misuse in the past 12 months of any of the specific pain relievers that were used in the past 12 months, where misuse was defined as use "in any way not directed by a doctor" ▪ First misuse of specific pain relievers that were misused in the past 12 months ▪ Lifetime misuse of pain relievers, if no past year misuse was reported ▪ Initiation of misuse of all pain relievers in the past 12 months ▪ Misuse of any pain reliever in the past 30 days, if misuse of specific pain relievers in the past 12 months was reported ▪ Additional questions for misusers in the past 30 days ▪ Additional questions for misusers in the past 12 months
Tranquilizers Main Section	▪ Misuse in the past 12 months of any of the specific tranquilizers that were used in the past 12 months, where misuse was defined as use "in any way not directed by a doctor" ▪ First misuse of specific tranquilizers that were misused in the past 12 months ▪ Lifetime misuse of tranquilizers, if no past year misuse was reported ▪ Initiation of misuse of all tranquilizers in the past 12 months ▪ Misuse of any tranquilizer in the past 30 days, if misuse of specific tranquilizers in the past 12 months was reported ▪ Additional questions for misusers in the past 30 days ▪ Additional questions for misusers in the past 12 months
Stimulants Main Section	▪ Misuse in the past 12 months of any of the specific stimulants that were used in the past 12 months, where misuse was defined as use "in any way not directed by a doctor" ▪ First misuse of specific stimulants that were misused in the past 12 months ▪ Lifetime misuse of stimulants, if no past year misuse was reported ▪ Initiation of misuse of all stimulants in the past 12 months ▪ Misuse of any stimulant in the past 30 days, if misuse of specific stimulants in the past 12 months was reported ▪ Additional questions for misusers in the past 30 days ▪ Additional questions for misusers in the past 12 months, including injection of stimulants

Table 1.2 **Content of Initial Substance Use Sections in the 2023 NSDUH Interview (continued)**

Section	Content
Sedatives Main Section	▪ Misuse in the past 12 months of any of the specific sedatives that were used in the past 12 months, where misuse was defined as use “in any way not directed by a doctor” ▪ First misuse of specific sedatives that were misused in the past 12 months ▪ Lifetime misuse of sedatives, if no past year misuse was reported ▪ Initiation of misuse of all sedatives in the past 12 months ▪ Misuse of any sedative in the past 30 days, if misuse of specific sedatives in the past 12 months was reported ▪ Additional questions for misusers in the past 30 days ▪ Additional questions for misusers in the past 12 months

Questions following this initial set of substance use questions asked about a variety of sensitive topics related to substance use and mental health issues ([Table 1.3](#)). These topics included (but were not limited to) injection drug use, perceived risks of substance use, substance use disorders (SUDs), arrests, treatment for substance use problems, pregnancy, mental illness, the utilization of mental health services, further details for the consumption of alcohol, and emerging issues related to substance use or mental health.

Additional demographic questions followed these sections on substance use, general health issues, and mental health issues ([Table 1.3](#)). These demographic questions asked about topics such as immigration, current school enrollment, employment and workplace issues, sexual attraction and sexual identity, English language proficiency, disabilities, and military families.

Once in-person respondents completed the ACASI portion of the interview, they returned the laptop computer to the interviewer. Remaining in-person interview questions were administered through CAPI and addressed household characteristics, health insurance coverage, and income ([Table 1.3](#)). The corresponding questions were modified as needed to be self-administered for web respondents.

Table 1.3 **Content of Sections on Remaining Topics in the 2023 NSDUH Interview**

Section	Content
Special Drugs	▪ Use of gamma hydroxybutyrate (GHB) or the use of cough and cold medications to get high ▪ Routes of heroin administration (i.e., smoking, sniffing, injection) ▪ Injection of cocaine, methamphetamine, or other drugs ▪ General injection use behaviors
Risk/Availability	▪ Perceived risk of harm associated with use of cigarettes, alcohol, or specific illicit drugs ▪ Perceived availability of illicit drugs ▪ Miscellaneous risk behaviors (e.g., seatbelt use)
Blunts	▪ Use of cigars with marijuana in them (i.e., “blunts”)
Substance Use Disorder	▪ Nicotine (i.e., cigarette) dependence ▪ Symptoms of substance use disorders for alcohol or drugs (marijuana, cocaine, heroin, hallucinogens, inhalants, methamphetamine, and any use of prescription drugs)

Table 1.3 **Content of Sections on Remaining Topics in the 2023 NSDUH Interview (continued)**

Section	Content
Special Topics	- Arrests in the lifetime and past 12 months - Probation and parole status, past 12 months - Driving under the influence of alcohol, marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine - Knowledge of state laws for marijuana possession
Prior Substance Use	- Use of marijuana, cigarettes, alcohol, or cocaine more than 12 months ago but within the past 24 months - Last use of substances included in the drug use section (see Table 1.2) if not used in the past 30 days
Alcohol and Drug Treatment	- Treatment for the use of alcohol or drugs (i.e., not counting tobacco) in the past 12 months - Perceived unmet need for treatment for use of alcohol or drugs - Barriers to receiving treatment for use of alcohol or drugs
Health	- Pregnancy status of females assigned at birth aged 12 to 44 - Height and weight - Use of health services in the past 12 months, including emergency room visits, hospitalizations, outpatient medical visits, and telehealth appointments - Conversations about tobacco, alcohol, or illicit drug use with healthcare professionals in the past 12 months - Sexually transmitted diseases in the past 12 months - Lifetime history of specific medical conditions, age at first diagnosis, and miscellaneous issues
Mental Health Services Utilization	- Treatment for mental health, emotions, or behavior in the past 12 months - Perceived unmet need for treatment for mental health, emotions, or behavior - Barriers to receiving treatment for mental health, emotions, or behavior
Social Environment 18+	- Specific illegal behaviors regardless of arrest (e.g., selling illegal drugs) - Attitudes about marijuana use - Religious involvement and beliefs
Parenting Experiences 18+ and parent of adolescent aged 12-17 also selected for the survey	- Beliefs about whether their child has used tobacco, alcohol, or drugs - Talks with their child about substance use - Attitudes about drug education
Youth Experiences 12-17	- Risk and protective factors for substance use - Fighting and delinquent behaviors - Attitudes and perceptions of others' attitudes about substance use - Exposure to drug prevention messages - Religious involvement and beliefs - Suicidal thoughts and behavior
Mental Health 18+	- Psychological distress in the past 30 days or past 12 months - Impairment in carrying out activities because of psychological distress - Suicidal thoughts and behavior
Adult Depression 18+	- Symptoms of depression in the lifetime and past 12 months - Impairment in carrying out activities because of depression symptoms - Treatment for depression

Table 1.3 **Content of Sections on Remaining Topics in the 2023 NSDUH Interview**
(continued)

Section	Content
Adolescent Depression 12-17	▪ Age-appropriate questions for symptoms of depression in the lifetime and past 12 months ▪ Age-appropriate questions for impairment in carrying out activities because of depression symptoms ▪ Treatment for depression
Consumption of Alcohol	▪ Number of drinks on the last occasion of alcohol use in the past 30 days ▪ Underage alcohol use ▪ Alcohol use in combination with illicit drugs ▪ Initiation of binge alcohol use
Emerging Issues	▪ Perceived recovery from problems with substance use or mental health (administered only to adults) ▪ Use of kratom, vaping of only flavoring, synthetic marijuana, synthetic stimulants, or illegally made fentanyl
Market Information for Marijuana	▪ Means by which respondents who used marijuana in the past 12 months obtained marijuana ▪ Price/value information for marijuana that was purchased ▪ Types of marijuana that were purchased ▪ Amount of marijuana purchased in joints or loose form
COVID-19	▪ Perceived effects of the COVID-19 pandemic on mental health, substance use, and financial situation ▪ Perceived effects of the COVID-19 pandemic on access to mental health treatment, substance use treatment, and medical care
Household Roster ¹	▪ Number of people living in the household, including the respondent ▪ Relationships of members of the household to one another
Proxy Information ¹	▪ Determination of whether another family member (a “proxy”) is better able to answer questions about the respondent’s health insurance and income ▪ “Hand-off” from the respondent to the proxy (if applicable)
Health Insurance ^{1,2}	▪ Current coverage by specific types of health insurance (if any) ▪ Periods without health insurance coverage ▪ Reasons for not having health insurance coverage
Income ^{1,2}	▪ Specific sources of income and related benefits ▪ Total personal income ▪ Total family income ▪ Miscellaneous questions on cell phone coverage

COVID-19 = coronavirus disease 2019.

¹ Questions are interviewer administered for in-person respondents.

² Questions are completed by the respondent or a proxy for the respondent.

1.2 Overall Data Quality Issues with CAI

The use of CAI (computer format) in person on laptop computers or via the web is designed to increase the quality of NSDUH data. For example, the programming logic that is embedded in CAI limited the opportunity for respondents to provide inconsistent answers. For example, when respondents are allowed to choose one or more responses from a list of prescription drugs they used in the past 12 months or to enter a response of “95” to indicate they have not used any of those prescription drugs from the list, the CAI program does not allow respondents to choose

items from the list and enter “95.” The occurrence of inconsistent data is further reduced by consistency checks and “hard errors,” built into the CAI program that prompt respondents to resolve inconsistencies that occurred between related items.³

However, the use of CAI did not completely eliminate data problems. For example, variables still could have missing data because CAI respondents had the option of entering a response of “don’t know” or “refused” when answering a given item. Similarly, even though consistency checks were designed to reduce inconsistent reporting, the CAI program was not equipped to address every possible inconsistent report that a respondent could make. Certain other quality issues could also be present in the CAI data. As discussed in Section 2.3.2, for example, the lack of direct interviewer monitoring could allow some respondents to use the computer keyboard to enter nonsensical patterns of answers for reasons such as accidental errors or disinterest such as keying responses of only 1 (or 11 or 111).

In addition, if respondents broke off the interview prior to reaching the end of the questionnaire, their data could still be used if they met the usable interview criteria (Section 2.2.3). Because interviewers were not present to keep web respondents engaged throughout the NSDUH interview (which had an average administration time of 60 minutes), web respondents broke off the interview at a higher rate than in-person respondents.

1.3 Organization of This Report

[Table 1.4](#) describes the content in the chapters of this report, and [Table 1.5](#) describes the content in the appendices. It is recommended that all readers first review Chapters 2 and 3 because the information in these two chapters helps set appropriate context for the more specific editing and imputation procedures that were employed on the variable sets presented in Chapters 4 through 14.

Table 1.4 Content of Chapters

Chapter	Content
1	Provides an overview of the NSDUH questionnaire; data quality issues; organization of the Editing and Imputation Report, which focuses on the editing and imputation procedures that were applied to respondent data; and survey or procedural changes from the previous year.
2	Discusses the preliminary coding and processing of unedited data as well as the general principles of logical editing.
3	Describes the predictive mean neighborhood (PMN) and modified predictive mean neighborhood (modPMN) imputation methodologies, which are used to impute NSDUH data.
4	Addresses the editing and imputation procedures that were applied to the front-end demographic variables.
5	Addresses the editing and imputation procedures that were applied to the back-end demographic variables.

³ When respondents are presented with a hard error message because of inconsistent answers, they cannot proceed further with the interview until they revise their answers to make them consistent. For other consistency checks, respondents are given an opportunity to revise their answers to make them consistent but can proceed with the interview regardless of whether they have resolved the inconsistency.

Table 1.4 **Content of Chapters (continued)**

Chapter	Content
6	Discusses the editing and imputation procedures for the substance use variables that form the basis for published estimates of substance use and misuse in NSDUH reports and tables.
7	Addresses the editing procedures for the special drugs and substance use disorder sections of the interview and the imputation procedures for relevant variables from the special drugs and substance use disorder sections, as well as the imputation procedures for cigarette dependence, which differ from the procedures used for other drug variables.
8	Discusses the editing procedures for additional sections on substance use and the imputation procedures for relevant alcohol and drug use treatment variables.
9	Discusses the editing procedures for sections on risk and protective factors for substance use. ¹
10	Discusses the editing procedures for sections on physical and mental health and the imputation procedures for relevant variables related to mental health services and adult mental health.
11	Describes the edits that were applied to the household roster, the creation of imputation-revised versions of the roster-derived household composition variables, and the creation of respondent-level variables with individual roster information.
12	Summarizes the editing and imputation procedures that were applied to the income variables.
13	Describes the procedures for logical editing and imputation of missing values in the health insurance variables.
14	Describes the editing and imputation processing of the pair relationship variable and related household count variables.

¹ The variables that are discussed in this chapter did not undergo imputation.

Table 1.5 **Content of Appendices**

Appendix	Content
A	Contains information about sampling and household-level and person-level eligibility and the completeness criteria used to construct the household-level and person-level files.
B	Provides a summary of data issues in different sections of the interview and the specific edits that were applied in response to these issues.
C	Provides details on the handling of "OTHER, Specify" responses to the race and Hispanic/Latino group questions so that the data could be summarized in a meaningful way.
D	Lists the covariates in each of the imputation models and includes (1) the starting list of covariates for each model and (2) descriptions of each level and identification of the reference level for categorical covariates.
E	Provides details on each final hot-deck step, or each collocated stochastic imputation step for categorical variables imputed using the modified predictive mean neighborhood (modPMN) imputation methodology, in the imputation procedures.
F	Summarizes the quality control measures used in the imputation procedures.
G	Lists a majority of the edit checks applied to the household roster.
H	Defines the rules for determining pair relationships.
I	Describes the conditions used for reconciling differing multiplicity counts between pair members.
J	Describes the conditions used for reconciling differing household-level person counts between pair members.

1.4 Changes from the 2022 Survey to the 2023 Survey

This section summarizes the changes in editing and imputation procedures implemented for NSDUH between the 2022 and 2023 survey years. It is intended for analysts who have had previous experience with the data. Those with little or no prior experience with NSDUH data are encouraged to review Chapters 2 and 3 before reviewing this section.

Section 1.4.1 describes changes to the 2023 questionnaire and includes, where relevant, a brief discussion of the effects on editing and imputation procedures for 2023. Starting with Chapter 4, the effects of the questionnaire changes outlined in Section 1.4.1 on editing and imputation procedures are discussed in greater detail as they relate to the relevant sections of the interview. Section 1.4.2 lists a few minor corrections and improvements to the editing and imputation procedures that also had little impact on the final edited or imputation-revised variables. The changes described in this chapter were implemented during the normal editing and imputation process.

1.4.1 Changes to Accommodate Modifications to the Questionnaire

The 2023 questionnaire included a few changes relative to the 2022 survey. To accommodate all questionnaire changes in 2023, existing SAS® programs for editing these variables were updated or replaced accordingly and are detailed in the subsections that follow. Where applicable, the changes to the 2023 questionnaire that required updates to the imputation procedures are also discussed.

1.4.1.1 New Self-Administered Front-End Demographics Section

A new self-administered front-end demographics section was included in the questionnaire. Questions about sex, Hispanic origin, race, and marital status were moved to this new section from the core demographics section, which was interviewer administered for in-person respondents. This change did not affect the editing or imputation procedures, and the corresponding edited and imputed variable names remained the same as in previous years.

1.4.1.2 Changes to the Sex Question

The question about respondents' sex was revised to ask about sex assigned at birth instead of asking whether respondents were male or female. In addition, respondents were given the opportunity to confirm that their response was correct and change the response, if applicable. Similar to the previous version of the question, respondents did not have the option of refusing to answer the question or indicating that they did not know the answer. The final variable, IRSEX, remained the same as in previous years.

1.4.1.3 New Questions about Gender Identity

New questions were included to ask about respondents' gender identity. Respondents could "enter all that apply" from the following response options to describe their current gender: male, female, transgender, and I use a different term. Respondents who indicated using a different term were asked in a follow-up question to enter the term they use to describe

their gender. Respondents were also asked to confirm that their responses to the gender questions were correct and were given the opportunity to change their original answer if they indicated that the answers were not correct. New edited variables were created corresponding to these new questions. These variables did not undergo imputation.

1.4.1.4 Changes to Sexual Attraction and Sexual Identity Variables

The questions about sexual attraction and sexual identity were revised to be administered to all respondents rather than to only adult respondents. The question about sexual identity was also revised for simplicity and to be more inclusive. Respondents were able to indicate using a different term for sexual identity as one of the response options, and then asked to enter that term in a new follow-up question. Edited variables corresponding to revised questions were assigned new variable names in 2023. A new edited variable was created corresponding to the new follow-up question. These variables did not undergo imputation.

1.4.1.5 Changes to Substance Dependence and Abuse

The substance dependence and abuse section was renamed the substance use disorder section for consistency with the questions that have been included since 2021 to estimate substance use disorder based on criteria in the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5; American Psychiatric Association, 2013). There were no changes in 2023 to the editing or imputation procedures corresponding to the questions in this section.

1.4.1.6 Removal of Medication-Assisted Treatment Questions in Emerging Issues Questionnaire Section

Questions that measured medication-assisted treatment (MAT) for alcohol and opioid use were removed from the emerging issues section. Since 2022, NSDUH has used MAT questions that were added to the alcohol and drug treatment section to estimate the receipt of MAT. The edited variables corresponding to these MAT questions from the emerging issues section are no longer available. The edited and imputed variables corresponding to the MAT questions in the alcohol and drug section were not impacted by the removal of the questions from the emerging issues section.

1.4.2 Changes Involving Minor Corrections and Enhancements

One change was made to the 2023 imputation procedures that involved a minor enhancement. In 2022, new questions were added to the marijuana section of the questionnaire that asked about the ways in which respondents used marijuana or cannabis products in the past 12 months and past 30 days. For the 2022 survey, only the variables related to vaping marijuana in the past 12 months and past 30 days were statistically imputed using the modified predictive mean neighborhood (modPMN) imputation method. Beginning in 2023, missing data in all variables related to the ways in which respondents used marijuana or cannabis products in the past 12 months and past 30 days were statistically imputed. The imputation method for 2023 was also changed from modPMN to predictive mean neighborhood (PMN). This change allows values from a single donor to be used to replace all missing values in a nonrespondent's record and reduces the amount of time required to process the marijuana ways-of-use variables. The

revised imputation procedures were also applied to the 2022 data, and the imputed variables were back delivered to the 2022 NSDUH analytic file, including reimputed versions of the past month and past year vaping variables. The estimates for past month and past year marijuana vaping reported in the *2022 National Survey on Drug Use and Health: Editing and Imputation Evaluation Report* (Center for Behavioral Health Statistics and Quality, 2023c), *Results from the 2022 National Survey on Drug Use and Health: Detailed Tables* (Center for Behavioral Health Statistics and Quality, 2023b), and *Key Substance Use and Mental Health Indicators in the United States: Results from the 2022 National Survey on Drug Use and Health* (Center for Behavioral Health Statistics and Quality, 2023a) used the original imputed data, that is, those imputed using modPMN. The 2023 Detailed Tables (Center for Behavioral Health Statistics and Quality, 2024c) include updated 2022 estimates for past month and past year marijuana vaping, which use the reimputed versions of these variables. [Table 1.6](#) shows the 2022 estimates for marijuana vaping using the original imputed data and the reimputed data, as reported in the 2022 and 2023 Detailed Tables.

Table 1.6 Comparison of Imputed Marijuana Vaping Estimates from the 2022 and 2023 Detailed Tables

Variable	Original 2022 Imputed Estimate	Reimputed 2022 Estimate
Marijuana Vaping		
Past Month ¹	34.7%	34.6%
Past Year ²	36.7%	36.6%

¹ Among past month marijuana users aged 12 or older.

² Among past year marijuana users aged 12 or older.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2022 and 2023.

2. Procedures and General Principles for Editing the NSDUH Data

2.1 Introduction

This chapter provides an overview of the procedures and general principles for editing the computer-assisted interviewing (CAI) data from the National Survey on Drug Use and Health (NSDUH). Unless otherwise specified, these procedures and principles applied to in-person and web interview data in 2023 because the basic content of questions in 2023 did not differ between these data collection modes.

Logical editing typically uses data from elsewhere within the same respondent's record in a deterministic manner to reduce the occurrence of missing or inexact data or to resolve inconsistencies between related variables.⁴ In contrast, statistical imputation procedures (see Chapter 3) apply probabilistic (stochastic) statistical methods to identify another respondent (i.e., a "donor") whose data are used to replace (1) missing values in the "recipient" respondent's record with nonmissing values from the donor or (2) inexact or nonspecific responses in the recipient respondent's record with more specific information from the donor.

As an example of inexact or nonspecific data, the CAI logic for the marijuana section required respondents to report that they used marijuana at least once in their lifetime in order to be asked when they last used it. However, respondents could report lifetime use of marijuana but not know or refuse to report the exact period when they last used it. In this case, the edited variable for most recent use of marijuana (MJREC) will be set to a nonspecific value (e.g., 9 = Used at some point in the lifetime logically assigned).

Section 2.2 describes the procedures for initially processing NSDUH interview data to prepare the data for further data processing. Coding of "OTHER, Specify" data and investigations of patterned responses are described in Section 2.3.⁵ Additional principles of data processing and editing are discussed in Section 2.4.

2.2 Initial Processing of NSDUH Interview Data

After in-person interview data were collected, the data were submitted to RTI as ASCII files. Web-based data were captured directly at RTI. Daily SAS[®] datasets were created from these

⁴ There are limited situations where data from outside the respondent's record is used in logical editing. For example, some editing procedures involve data from the screener where an eligible member of the dwelling unit reports basic information about all members of the dwelling unit. Further, in situations where two members of the same dwelling unit are selected for the survey and complete the interview, data from the second respondent's record may be used in logical editing of some variables in the first respondent's record, or vice versa. This procedure allows use of information from both respondents to determine relationships among household members.

⁵ "OTHER, Specify" data refer to the alphanumeric answers that respondents or interviewers type and the numeric codes associated with those responses.

in-person and web data. This daily processing included the following activities as part of the creation of initial unedited interview data files:

- assignment of standard NSDUH missing data codes (Section 2.2.1),
- remapping of responses to “enter all that apply” questions (Section 2.2.2), and
- identification of “usable” interviews (Section 2.2.3).

Each day’s processed SAS dataset was merged with the interview data that had been received to date. This process continued until the end of the quarter (e.g., January through March = Quarter 1) when a cumulative data file (Section 2.2.4) was produced that contained all interviews that had been received from that quarter.

2.2.1 Assignment of Standard NSDUH Missing Data Codes

Following receipt of the unprocessed in-person and web-based interview data, a key activity in the initial processing of the data involved the assignment of standard NSDUH missing data codes. The Blaise⁶ program for the CAI instrument uses a code of 8 (or 98 or 998, etc.) to denote responses of “refused” (REF), a code of 9 (or 99 or 999, etc.) to denote responses of “don’t know” (DK), and a blank to denote questions that were not answered because of the CAI skip logic or because respondents broke off the interview before completing the survey. In final NSDUH data files, however, the codes for missing data that were supplied by the Blaise program were replaced with the standard NSDUH codes for DK, REF, and “blank.”

The following standard codes⁷ for missing data were relevant to both the initially processed unedited data and the edited data:

- 94 (or 994 or 9994, etc.) = DON’T KNOW (DK);
- 97 (or 997 or 9997, etc.) = REFUSED (REF); and
- 98 (or 998 or 9998, etc.) = BLANK (i.e., nonresponse).

Assignment of codes as part of the editing procedures to indicate when a question was legitimately skipped is discussed in Section 2.4.2.

2.2.2 Remapping of Responses to “Enter All That Apply” Questions

A second activity associated with the initial processing of unedited in-person and web-based interview data involved the remapping of responses to “enter all that apply” type questions.

⁶ Blaise is the computer program within the CAI instrument that was used to direct the respondent and interviewer through the questionnaire.

⁷ Codes for missing data in most unedited variables were two digits in length (i.e., 94, 97, or 98). For some variables, however, values in the 90s were part of the allowable range of valid responses. For example, respondents could report the total number of days they used a substance in the past 12 months, with an allowable range of 1 to 366 days (to take leap years into account). In this situation, respondents could have used a substance on 94, 97, or 98 days. Therefore, for variables that indicate the total number of days that respondents used a substance in the past 12 months, the codes for missing data were three digits in length (i.e., 994, 997, or 998).

“Enter all that apply” questions allowed respondents to choose as many responses from a given list as applied.

The CAI program captured information from these “enter all that apply” questions as separate variables in the order that respondents selected their answer choices. For example, the CAI data that were initially received included five separate variables for the specific pain relievers listed in question PR01 to accommodate reports of any use in the past 12 months of these pain relievers, or to allow respondents to choose a response of “95” to indicate that they did not use any of the past relievers listed in PR01 in the past 12 months.⁸ Consequently, an “enter all that apply” variable in the received data could have a different meaning across respondents, depending on how respondents answered the question. For example, if respondents reported that they had not used any of the pain relievers listed in PR01 in the past 12 months, then a response of 95 would be captured in the first variable from the received data, and the remaining “unused” variables from question PR01 would be blank. If other respondents reported the use of generic hydrocodone and no other pain relievers from PR01 in the past 12 months, then a response of 5 (corresponding to the response option for hydrocodone) would be captured as the first variable in the received data for PR01. If a third group of respondents chose Vicodin® and hydrocodone as their first and second responses from the list, then the first variable from PR01 would be a response of 1 (corresponding to the response option for Vicodin®), the second would be a response of 5 (corresponding to the response option for generic hydrocodone), and the remaining variables would be blank.

For individual “enter all that apply” variables to have the same meaning for all respondents, responses were remapped according to the order of the response choices that respondents saw in the questionnaire. For example, six discrete unedited variables associated with question PR01 were created as part of the remapping process to capture all reports of any use of Vicodin® in the past 12 months in the variable PR011 (i.e., the first variable in the series); reports of any use of Lortab® in the variable PR012; reports of any use of Norco® in the variable PR013; reports of any use of Zohydro® ER in the variable PR014; reports of any use of generic hydrocodone in the variable PR015; and reports of nonuse of any prescription drug in the list in the past 12 months (i.e., responses of 95) in the variable PR016. If respondents did not report the use of a specific drug from the list of pain relievers in PR01, then the remapped variable for that pain reliever was assigned a code of 98 (i.e., blank). Thus, if respondents chose an answer of 95, then PR016 was given a code of 95 and the remaining variables for PR011 through PR015 were assigned codes of 98. If respondents reported the use of at least one drug from the list in PR01, then PR016 received a code of 98.

In addition to choosing one or more applicable responses from a list, respondents could answer “don’t know” or “refused” as their first response to these “enter all that apply” types of questions. In situations where respondents answered “don’t know,” it would be reasonable to infer that the respondents did not know which particular item on the list applied to them. In this situation, a code of “don’t know” was propagated to each of the remapped “enter all that apply”

⁸ The CAI program does not allow respondents to report the use of one or more specific prescription drugs in a list and also report no use of any of the listed prescription drugs in the past 12 months.

variables as part of the daily processing of the received data. Similarly, if respondents refused to answer question PR01, a refusal code was propagated to all of the relevant remapped variables as part of the daily processing of the received data.

2.2.3 Identification of Usable Interviews

The third step in the preliminary processing of in-person and web-based data determined whether a received interview met the minimum item response requirements to be retained for weighting and further analysis (subsequently referred to as “usable” data).⁹ These rules were designed to eliminate interviews with unacceptable levels of item nonresponse (i.e., missing data). The final dataset would only use data from interviews with lower levels of missing data and therefore would reduce the amount of statistical imputation needed for any given record. In 2023, about 1 percent of all received interviews (approximately 700) did not meet the minimum requirements to be considered usable. For in-person interviews, about 0.3 percent did not meet minimum usability requirements. For interviews that were received from web respondents, about 2.3 percent did not meet minimum usability requirements.

The usability criteria established for NSDUH are based on the completeness of information respondents provide about their lifetime use or nonuse of different substances. In NSDUH, respondents are asked more detailed questions about different substances only if they reported lifetime use of that substance or category of substances on an initial question. These questions are subsequently referred to as gate questions because reporting of lifetime use opens the “gate” to the administration of additional questions. The “gate” is closed if none of the responses to the related questions explicitly indicate use.

For a NSDUH record to be considered usable, regardless of whether data collection occurred in person or via the web, both of the following requirements must be met:

1. The lifetime cigarette question (i.e., “Have you ever smoked part or all of a cigarette?”) had to have been answered as “yes” or “no.” This requirement was set so that lifetime use or nonuse would be fully defined for cigarette use.
2. Responses to questions on at least nine (9) of the following additional substances had to contain information about use or nonuse: (a) smokeless tobacco, (b) cigars,¹⁰

⁹ The historical background and considerations for establishing usable interview rules for the CAI data are discussed in a methodological chapter on editing the 1999 CAI data (Krouil & Myers, 2002). For more information about the development of the usable interview requirements for the redesigned NSDUH questionnaire, see the *National Survey on Drug Use and Health: 2012 Questionnaire Field Test Final Report* (Center for Behavioral Health Statistics and Quality, 2014).

¹⁰ Although NSDUH respondents were also asked about nicotine vaping and pipe tobacco, these measures were not included in the usability criteria. Pipe tobacco use was not included because there was only one other question about pipe tobacco in addition to the lifetime pipe tobacco use question. Although the nicotine vaping questions appeared in the same section with questions on tobacco products that were included in the usability criteria, they were moved from the emerging issues section in 2022. Including nicotine vaping in the criteria would have a minimal effect on the outcomes for usability, and keeping the criteria the same as in prior years made sense.

(c) alcohol, (d) marijuana, (e) cocaine (in any form),¹¹ (f) heroin, (g) hallucinogens, (h) inhalants, (i) methamphetamine, (j) prescription pain relievers, (k) prescription tranquilizers, (l) prescription stimulants (i.e., independent of methamphetamine), and (m) prescription sedatives.

The usability criterion for smokeless tobacco through heroin and for methamphetamine (i.e., the drugs listed above as items a to f and item i) was that lifetime use or nonuse must be determined from the single gate question. For hallucinogens and inhalants (items g and h above), which had multiple “gate questions,” at least one gate question in the series for that section was required to have an answer of “yes” or “no.”¹² Any of the following met the usability criteria for prescription drugs (items j to m above):

- use in the past 12 months of at least one specific prescription drug in a category (e.g., pain relievers) was reported;
- lifetime use or nonuse of any prescription drug in the category was reported; or
- nonuse of all specific prescription drugs in the past 12 months was reported, regardless of whether lifetime use or nonuse could be determined.

The interview also included follow-up probes for respondents who initially refused to answer a gate question or an entire series of gate questions. Follow-up probes were included in the following sections that were relevant to the usable interview criteria: cigarettes, smokeless tobacco, cigars, alcohol, marijuana, cocaine, heroin, specific hallucinogens (i.e., LSD, PCP, Ecstasy/“Molly” [MDMA]), any use of inhalants, and methamphetamine. If respondents changed their initial refusal to a response of “yes” or “no,” they were considered to have provided usable data to that drug’s gate information.

In addition, the CAI program terminated interviews during the initial demographics questions if respondents were ineligible for the survey (i.e., younger than age 12 or on active duty in the U.S. military) or if respondents did not provide their birthdate or age. Also, interviews were not able to proceed if certain initial questions, such as sex assigned at birth and state of residence, were not answered. Data from these ineligible or incomplete records did not meet the usable interview criteria because the interviews were terminated before the respondents were asked the initial cigarette use question.

Meeting the usable interview criteria was a necessary, but not sufficient, requirement for an interview to be considered a final respondent interview. Respondents who had sufficient data to meet the usable interview criteria could still be treated as nonrespondents if their interview data

¹¹ Crack cocaine was not included in the usable interview criteria because the logic for asking about crack cocaine was dependent upon the respondent having answered the lifetime cocaine question as “yes.”

¹² For hallucinogens and inhalants, the logic for asking more detailed questions about use of that category of drugs was based on respondents’ answers to multiple “yes/no” questions about the lifetime use or nonuse of specific drugs within that category (e.g., lifetime use or nonuse of the following specific hallucinogens: lysergic acid diethylamide [LSD]; phencyclidine [PCP]; peyote; mescaline; psilocybin; Ecstasy or “Molly” [3,4-methylenedioxymethamphetamine (MDMA)]; ketamine; DMT, AMT, or “Foxy”; *Salvia divinorum*; or “any other” hallucinogen).

appeared to be of poor data quality, as evidenced by potential response pattern problems (Section 2.3.2).

2.2.4 Creation of Cumulative Quarterly Unedited Data Files

Following daily processing of the data, each day's SAS dataset was appended to the cumulative data that had been received up to that point in the quarter. At the end of the quarter, a complete data file was produced that contained all interviews that had been received during the quarter. Each quarterly data file then underwent additional cleaning and processing to modify or correct errors (e.g., erroneous ID entries by staff for in-person interviews) before the data were ready for editing. The cumulative cleaned-up (but otherwise unedited) SAS dataset from the entire survey year served as the starting point for the final logical editing procedures that are described in Chapters 4 through 14.¹³

2.3 Coding of "OTHER, Specify" Data and Investigation of Response Patterns

Following the initial procedures that were described in Section 2.2, the following activities took place prior to the creation of edited variables:

- coding of "OTHER, Specify" data (Section 2.3.1); and
- investigation of response patterns in records that otherwise met the usable interview criteria (Section 2.3.2).

To facilitate timely review and processing, coding of "OTHER, Specify" data and review of the data for patterned responses could occur prior to creation of the cumulative quarterly unedited data files (Section 2.2.4).

Codes to denote "bad data" (i.e., to indicate an inconsistency or some other problem in the original data) could be assigned to variables during any of these activities following the processing of received interviews. The following codes were assigned to denote "bad data": 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned. Codes for bad data were treated as missing values.

2.3.1 Coding of "OTHER, Specify" Data

Some questions in NSDUH were open-ended and asked respondents to type in a response. Coding is the activity that took alphanumeric (text) answers that respondents or interviewers had typed for the open-ended questions (e.g., specific other drugs that were used or misused, specific other places respondents stayed overnight or longer for treatment for alcohol or drug

¹³ Some edits are run on preliminary data from the first two quarters of a given survey year to identify any updates that need to be made to the programs for use with the full data from all four quarters. Running the edits on data from the first two quarters is particularly useful for testing the programs for any sections of the CAI instrument that are new or have changed since the previous survey year.

use) and converted them into numeric codes. These alphanumeric answers and the numeric codes resulting from them are referred to as “OTHER, Specify” data.

Coding of the “OTHER, Specify” variables was accomplished through computer-assisted procedures.¹⁴ “OTHER, Specify” responses were first converted to all capital letters because respondents could use different combinations of uppercase or lowercase characters to provide an otherwise identical response. If an exact match was found between what the respondent keyed and an existing entry in the data dictionary (e.g., “ALCOHOL”), the computer-assisted procedures assigned the appropriate numeric code (e.g., 807 for alcohol).¹⁵ The system could also accommodate commonly encountered misspellings (e.g., “ALCHOHOL”) that had previously been entered into the data dictionary.

Typed answers that the respondent provided that did not match an existing data dictionary entry were reviewed by analysts on a flow basis during the quarter to determine whether an existing code should be assigned to the response or a new code should be created. Based on these decisions, new entries were added to the relevant dictionaries on a quarterly basis—including new dictionary entries corresponding to existing codes—for use in daily processing of “OTHER, Specify” data in subsequent quarters. Analysts could also decide not to add a particular response to the data dictionary, in which case the response would be output for review and coding on a case-by-case basis if it was reported in a future quarter or survey year. In addition, analysts reviewed the codes that were assigned through the computer-assisted process to verify that these codes were being assigned correctly. Over time, these procedures have reduced the turnaround time and burden on analysts for producing the coded “OTHER, Specify” data.

Particularly for other drugs (Section 2.3.1.1), which used the same data dictionaries for a number of “OTHER, Specify” variables, matching a written response to a numeric code in the data dictionary would result in that code being assigned no matter where respondents typed that answer. For example, a given entry in the data dictionary for drugs would be assigned the same code if it appeared in the hallucinogens section or in the pain relievers section.

As with the “enter all that apply” data that were discussed previously, respondents could answer the “OTHER, Specify” questions as “don’t know” or “refused,” which were then reassigned to the respective codes of 94 (or 994 or 9994) or 97 (or 997 or 9997), as described in Section 2.2.1. Respondents could also type in an equivalent response to “don’t know” (e.g., “no idea”) or “refused” (e.g., “too personal”). These equivalent responses were assigned the relevant codes for missing data (Section 2.2.1) as part of the coding procedures. For typed

¹⁴ A system has been in place since 2002 for the daily coding and processing of the “OTHER, Specify” variables for drugs. This system now encompasses the daily coding and processing of most “OTHER, Specify” variables from the CAI instrument that underwent assignment of numeric codes, except for the codes pertaining to race or ethnicity that are described in Section 2.3.1.2.

¹⁵ A data dictionary is a database of previous responses to a given category of related “OTHER, Specify” questions (e.g., illicit drugs, locations for the receipt of substance use treatment), including misspellings and slang terms (where applicable), that are used to code typed responses.

responses that were nonsensical or otherwise nonresponsive to the request to specify additional information, codes for bad data (Section 2.3) were assigned.

The remainder of the discussion in this section focuses on issues related to coding of the “OTHER, Specify” data according to the type of other information that was requested from respondents:

- other drugs (Section 2.3.1.1),
- other race or ethnicity (Section 2.3.1.2),
- other types and locations for substance use treatment and mental health treatment (Section 2.3.1.3), and
- additional “OTHER, Specify” data (Section 2.3.1.4).

Except for mentions of other drugs (Section 2.3.1.1), other ways that respondents used marijuana or cannabis products, and other ways respondents got marijuana or cannabis products (Section 2.3.1.4), “OTHER, Specify” data in the 2023 NSDUH were intended to capture a single “other” response (e.g., other outpatient location where respondents received mental health treatment in the past 12 months).¹⁶ If respondents typed in responses for which multiple codes could apply, the standard procedure was to assign a code to the first response that could be coded. Exceptions to this general procedure are noted for the coding of “OTHER, Specify” data.

2.3.1.1 Other Drugs

Respondents could specify other drugs as follows, according to the relevant questions they were asked:

- in the hallucinogens section, other hallucinogens they used in their lifetime;
- in the inhalants section, other inhalants they used in their lifetime;
- in the relevant sections for prescription drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), other prescription drugs they misused in the past 12 months;
- in the special drugs section, other drugs they used with a needle in their lifetime;¹⁷
- in the substance use disorder section, other drugs they used to get over or avoid having withdrawal symptoms for a particular drug; or
- in the alcohol or drug use treatment section, other drugs for which they received treatment as an inpatient or outpatient in the past 12 months.

¹⁶ “OTHER, Specify” questions for (1) use of specific drugs in the past 12 months to get over or avoid having withdrawal symptoms asked in the substance use disorder section and (2) receipt of treatment in the past 12 months for a specific drug asked about in the alcohol and drug use treatment section were designed to capture a single “other” response. However, similar to the remaining other drug variables, five coded variables were created for each question. This allowed more responses to be captured if respondents typed in multiple drugs for the single response.

¹⁷ Specific procedures for the coding of “OTHER, Specify” questions in the special drugs section are described in Section 7.2.1 for the use of other over-the-counter cough and cold medicines to get high in the past 12 months.

The basic “OTHER, Specify” coding scheme for drugs was as follows:

- Codes of 101-199 and 1001-1999 were reserved for prescription sedatives.
- Codes of 201-299 and 2001-2999 were reserved for prescription tranquilizers.
- Codes of 301-399 and 3001-3999 were reserved for prescription stimulants.
- Codes of 401-499 and 4001-4999 were reserved for prescription pain relievers.
- Codes of 501-599 and 5001-5999 were reserved for inhalants.
- Codes of 601-699 and 6001-6999 were reserved for hallucinogens.
- Codes of 701-799 were reserved for heroin and other opiates (e.g., opium).
- Codes of 801-899 and 8001-8999 were reserved as a residual category for tobacco products, alcohol, cannabis (e.g., marijuana), cocaine (including crack), and miscellaneous drugs that did not fall into any of the other categories mentioned previously (e.g., prescription drugs that were not sedatives, tranquilizers, stimulants, or pain relievers).
- Code 900 was reserved as a special code when respondents specified something that was not a drug (often in the sections for substance use disorder and alcohol and drug use treatment).

Prescription drugs were assigned codes in the 800 or 8000 series if they did not belong pharmacologically to any of the prescription drug categories in the NSDUH questionnaire, regardless of where respondents specified these drugs. For example, if a respondent specified an antidepressant as “some other tranquilizer,” then the antidepressant was coded in the 800 or 8000 series (rather than in the 200 or 2000 series) because it was not a tranquilizer.

2.3.1.1.1 Coding for Hallucinogens, Inhalants, Prescription Drugs, and Special Drugs

In the “OTHER, Specify” drug questions in the sections for hallucinogens, inhalants, prescription drugs, and special drugs, respondents could enter up to five responses (in five data entry fields) for other drugs. For the most part, respondents specified a total of only one drug in these questions or they specified only one drug in each field. Neither of these situations required any special handling of the data.

However, additional procedures were required to handle the following situations in the “OTHER, Specify” drug data for these sections:

1. Respondents specified more than one drug within a single data entry field, but they specified a total of five or fewer drugs across the five available fields.
2. Respondents specified a total of more than five substances across the five available data entry fields (i.e., by definition, they specified multiple drugs within one or more fields).

In the first situation, codes for the additional drugs were moved to the next unused “slot(s)” (i.e., variable[s]). For example, if three drugs were specified in total, with two of them being specified in the first field and the third being specified in the second field, the additional code from the first field was moved to the third “OTHER, Specify” variable. Consequently, the

responses in the individual “OTHER, Specify” drug variables did not always correspond to the order in which respondents reported use of these drugs.

When respondents specified more than five substances in the available fields, duplicate mentions of drugs were identified and removed. Duplicate mentions could include the same drug being mentioned more than once in the “OTHER, Specify” data or a drug being reported in the gate questions for a given section and also being reported in the “OTHER, Specify” data. For example, if question LS01a for LSD was answered as “yes” and LSD also was reported in the “OTHER, Specify” data for hallucinogens, then the duplicate code for LSD was not kept in the “OTHER, Specify” data for hallucinogens to allow codes to be used for other drugs. If more than five unique mentions of drugs remained after any redundant mentions had been eliminated, further priority in retaining responses in the final drug codes was given to (1) mentions of illegal or prescription drugs, as opposed to “over-the-counter” drugs that are available in the United States without a prescription and (2) mentions that were relevant to the category of interest (e.g., in the hallucinogens section, giving priority to mentions of hallucinogens over mentions of drugs that were not classified as hallucinogens). Because the “OTHER, Specify” questions in the special drugs section for injection of other drugs applied to substances in any category, this second criterion regarding the relevant category of interest was not used in deciding which codes to keep for special drugs.

In rare instances, these procedures still did not yield five or fewer unique drugs in the “OTHER, Specify” data for a given section. In these situations, the highest priority was given to retaining codes for drugs that had no approved medical use in the United States or were deemed to have a greater potential for abuse, and the lowest priority was given to retaining codes for drugs that were not classified as controlled substances.¹⁸ In the extremely rare situations where five or more unique drugs remained after all of these measures had been applied, the first five remaining “OTHER, Specify” codes were retained, and mentions of any additional drugs were dropped from the coded data that were available for further editing or analysis.

2.3.1.1.2 Coding for Substance Use Disorder and Alcohol and Drug Use Treatment

In the substance use disorder section, respondents were asked about their use of substances to get over or avoid having withdrawal symptoms for the substances that have a withdrawal criterion¹⁹ and that they used in the past 12 months.²⁰ Respondents who reported using “something else” were provided a single data entry field to report the other substance they used to get over or avoid withdrawal symptoms. Similarly, in the alcohol and drug use treatment section, respondents could report that they received inpatient or outpatient treatment

¹⁸ The drug scheduling classifications, [Schedules I through V](#), under the [Controlled Substances Act](#) were used in making these determinations.

¹⁹ These substances are alcohol, marijuana, cocaine, heroin, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription stimulants, and prescription sedatives. Hallucinogens and inhalants do not have a withdrawal criterion.

²⁰ The questions for alcohol use disorder and marijuana use disorder applied to respondents who used the respective substances on 6 or more days in the past 12 months.

for “some other drug.” These respondents were provided a single data entry field to report the other drug for which they received inpatient or outpatient treatment.

Even though respondents were provided with only a single data entry field, the coding procedures captured data for up to five substances, in case respondents reported multiple substances in the single data entry field. If respondents reported more than five substances, then the procedure that was described in Section 2.3.1.1.1 was followed.

2.3.1.2 Other Race or Ethnicity

In the front-end demographics section, NSDUH respondents were asked in question QD03 if they were of Hispanic/Latino origin and to report their race in question QD05 (see Section 4.2.7). Respondents who indicated that they were of Hispanic, Latino, or Spanish origin, that they were Native Hawaiian or Other Pacific Islander, or that they were Asian were subsequently asked to report which group best described them (QD04, QD05PI, and QD05ASIA, respectively). Respondents who considered that the categories presented to them for Hispanic/Latino origin, Pacific Islander ethnicity, or Asian ethnicity did not apply to them could report their “other” Hispanicity, Pacific Islander ethnicity, or Asian ethnicity. Respondents who selected “Don’t Know” for race could report their “other” racial group.

The computer-assisted coding procedures and use of data dictionaries that were described in Section 2.3.1 applied to these “OTHER, Specify” data for race and ethnicity. In most situations, new responses were just new misspellings of an already established category, such as a response of “Porto Rican” instead of “Puerto Rican.”

Regardless of the source (QD04, QD05, QD05PI, QD05ASIA) for these “OTHER, Specify” data, the responses were used in subsequent editing of Hispanicity and race to determine the respondents’ final Hispanicity and race (see Sections 4.2.7, 4.3.2, and 4.3.3). Thus, in coding the “OTHER, Specify” data, each response was assigned two codes: one for race and the other for Hispanicity.

2.3.1.3 Other Types and Locations for Substance Use Treatment or Mental Health Treatment

In the sections for alcohol and drug use treatment and mental health services utilization, respondents were asked whether they received treatment in some other place where they stayed overnight or longer (inpatient) and whether they received treatment in some other place as an outpatient. For each of these questions, respondents who reported receiving treatment in some other place were asked an “OTHER, Specify” question to describe the other place where they received treatment. For example, respondents who indicated receiving mental health treatment in some other place as an outpatient in question TXMHOPG were asked in question TXMHOPGO to describe the other place where they received outpatient mental health treatment.

Because treatment in some locations (e.g., a hospital) may be provided on an inpatient or outpatient basis, an important aspect in coding these “OTHER, Specify” data was determining whether respondents received the treatment as an inpatient or an outpatient. The following

coding assumptions were made in situations where either inpatient or outpatient treatment could be provided at a specified location:

- If respondents specified a location but did not designate whether they received treatment as an inpatient or an outpatient, then the classification was assumed to match the classification asked about in the question.
 - For example, respondents who specified “hospital” in question TXSBIPDO for substance use treatment in some other inpatient location were assumed to have received treatment at a hospital as an inpatient.
 - Similarly, respondents who specified “hospital” in question TXSBOPGO for substance use treatment in some other outpatient location were assumed to have received treatment at a hospital as an outpatient.
- If respondents specified a location and also specified that they received treatment in that location as an inpatient or an outpatient, then that response took precedence over the question in which they specified the location. For example, if respondents specified that the other location where they received substance use treatment as an outpatient could be interpreted to be in a hospital where they stayed overnight, then respondents were coded as having received this “outpatient” treatment at a hospital as an inpatient.

In addition, questions for substance use treatment and mental health treatment used “rehab” for substance use treatment facilities (e.g., a drug or alcohol rehab or treatment center as an outpatient) but not for mental health treatment facilities (e.g., a mental health treatment center as an outpatient). However, respondents also could interpret a rehabilitation facility to refer to a mental health treatment facility. Therefore, if respondents specified “rehab” by itself as some other location, it was not automatically assumed that respondents received treatment in a substance use treatment facility. For example, if respondents specified “rehab” as the other place where they stayed overnight to receive treatment for their alcohol or drug use, the coding procedures did not assume that respondents received inpatient substance use treatment in a residential drug or alcohol rehab or treatment center. In this example, respondents were assigned a special code for “Inpatient; facility type unspecified.”

A similar principle applied if respondents specified only “rehab” as the other place where they received treatment as an outpatient. For example, if respondents specified only “rehab” as the other place where they received treatment as an outpatient for their use of alcohol or drugs, then their response was assigned a special code for “Outpatient; facility type unspecified.”

2.3.1.4 Additional “OTHER, Specify” Data

There were three types of additional “OTHER, Specify” questions in the remaining questionnaire sections:

- those where the “OTHER, Specify” item was a follow-up to a lead question that typically was answered as “yes” or “no”;²¹
- those where the “OTHER, Specify” item was a follow-up to a response category for “other” in an “enter all that apply” question; and
- those where respondents did not get the opportunity to choose the “other” response (and specify something) if they already chose another category from the list.

Coding of these “OTHER, Specify” variables was performed according to the general principles described in Section 2.3.1. Otherwise, minimal additional decision making was involved in assigning codes to the responses for these variables. Descriptions of these variables are included here for completeness.

An example of the first type of “OTHER, Specify” question is question MJE01E2 in the market information for marijuana section (i.e., specify the other way that respondents got marijuana or cannabis products). Only those respondents who reported in question MJE01E that they got marijuana or cannabis products in some other way (i.e., question MJE01E was answered as “yes”) were routed to MJE01E2 and asked to report the other way. The first four “OTHER, Specify” codes corresponded to the response choices in questions MJE01A through MJE01D, in case respondents specified a way they got marijuana or cannabis products that corresponded to one of the previous ways they were asked about. For example, a code of 1 was assigned if respondents specified that the other way in which they got marijuana or cannabis products was by paying for or buying them. These codes were used in editing the variables corresponding to MJE01A through MJE01D.

Another example of this first type of “OTHER, Specify” question is question QD15 (“In what other country or U.S. territory were you born?”) in the back-end demographics section. For this “OTHER, Specify” question, respondents were routed to QD15 if they answered question QD14 (“Were you born in the United States?”) as “no.” The first set of “OTHER, Specify” codes corresponded to states in the United States, including a code if respondents specified that they were born in the United States, but they did not specify a state. If respondents specified that they were born in Alaska or Hawaii before they became states, they were coded as having been born in the respective states. Codes starting at 60 were reserved for U.S. territories. Codes starting at 100 were reserved for other countries and were grouped by continent. For example, codes in the 100s were used for respondents who specified that they were born in Europe, not counting situations in which respondents specified that they were born on a U.S. military base. A separate code was used for respondents who specified that they were born on a U.S. military base outside of the United States.

Question MJMDPYOA in the marijuana section (i.e., specify the other way respondents used marijuana or cannabis products in the past 12 months) is an example of the second type of “OTHER, Specify” question. In question MJMODEPY, respondents could report up to eight different ways that they used marijuana or cannabis products in the past 12 months, including

²¹ Depending on the nature of the lead question, either an affirmative or a negative response to the lead question could govern whether respondents were asked to specify something.

"some other way." Respondents were routed to MJMDPYOA if they reported using marijuana or cannabis products in some other way in MJMODEPY. Because MJMODEPY was an "enter all that apply" question, respondents could report using marijuana in some other way, along with any of the specific ways to use marijuana or cannabis products that were listed in the question (e.g., smoking). The first seven "OTHER, Specify" codes corresponded to the response choices in MJMODEPY, in case respondents specified a way they used marijuana or cannabis products in the past 12 months that corresponded to one of the previous ways they were asked about. For example, a code of 1 was assigned if respondents specified that the other way in which they used marijuana or cannabis products was by smoking. These codes were used in editing the separate "enter all that apply" variables corresponding to the first seven response choices in MJMODEPY.

An example of the third type of "OTHER, Specify" question is represented by question SD21SP in the special drugs sections (i.e., specify the other way respondents got the needle the last time they used a needle to inject drugs). Respondents were routed to question SD21SP if they selected the "I got the needle in some other way" option in question SD21. SD21 contained a list of five response options, including the last option for getting the needle in some other way. However, respondents were permitted to select only one of the five options. If respondents chose any response from the list of options in SD21 except for "other," they were *not* routed to the "OTHER, Specify" question. Data from SD21 and SD21SP were combined into a single final variable, GNNDGET. When respondents chose the other category in the lead question, but they specified something that was coded as a missing value (i.e., don't know, refused, bad data, or blank), then GNNDGET retained a code corresponding to other, as opposed to being assigned a missing value.

The edits applied to GNNDGET were designed to provide analysts with a standardized way to readily identify when it could be logically inferred that respondents should have chosen a given response option from the preceding question (i.e., rather than "other"). Codes of 1 to 4 applied to answers that respondents gave directly from question SD21 (e.g., 1 = Bought the needle from a pharmacy). Category 5 in SD21 was "Got the needle some other way."

Responses in the "OTHER, Specify" variables for GNNDGET that corresponded to existing response categories in their respective lead questions were coded starting with the number 21, with the coding proceeding in the order of the existing response categories. This procedure allowed for the possibility that lead questions could have 10 or more available choices and prevented overlap of "OTHER, Specify" codes with available responses. This procedure also offered analysts a standardized way to interpret these values. For example, if analysts wanted to treat codes of 21 to 24 in GNNDGET that came from "OTHER, Specify" data as being equivalent in an analysis to the corresponding codes of 1 to 4, the analyst could simply subtract 20 from any codes with values of 21 to 24 to recode these values to the corresponding codes of 1 to 4.

For similar reasons, the coding sequence for responses in these "OTHER, Specify" variables that did not correspond to responses from the list in the lead question resumed at number 41 (e.g., for GNNDGET, 41 = Given by/stolen from friend/acquaintance of friend/nonrelative). Although codes for responses that did not correspond to available choices from a lead question could have

resumed at a lower number—especially for SD21SP—this approach standardized the assignment of codes across these types of “OTHER, Specify” variables (i.e., new codes always started at 41), minimized the risk of overlap between codes for these two types of responses, and allowed flexibility if new response choices were added to a lead question in a future year.

2.3.2 Investigation of Response Patterns in Usable Interviews

The computer-assisted environment for self-administration of questions could encourage some respondents to use their computer keyboard or other input device (for web interviews) to enter nonsensical patterns of answers if they were not paying attention to questions or were not taking the interview seriously for other reasons. Also, respondents could miss that the meaning of a response had changed, such as if a response of 2 meant “no” in previous questions but had a different meaning in a new set of questions.

Thus, even if respondents had sufficient data for their interviews to be considered usable (Section 2.2.3), certain patterns of answers could call into question the validity of the data, either for the entire interview or for specific sections of the interview. [Table 2.1](#) provides a summary of questionnaire sections that are screened for response patterns. If any of the sections are flagged as having potential response patterns, all questionnaire sections within a category (i.e., nonprescription substances, prescription drugs, and questionnaire sections following the substance use sections) are examined further.

Table 2.1 Questionnaire Sections Screened in Data Diagnostic Review

Sections Screened in Nonprescription Substances Program	Sections Screened in Prescription Drugs Program	Sections Screened in Non-Substance Use Program ¹
Nicotine	Pain Relievers Screener	Alcohol and Drug Use Treatment
Alcohol	Tranquilizers Screener	Mental Health Services Utilization
Marijuana	Stimulants Screener	Youth Experiences
Cocaine	Sedatives Screener	Back-End Demographics ²
Crack	Pain Relievers Main	Education ²
Heroin	Tranquilizers Main	Employment ²
Hallucinogens	Stimulants Main	None
Inhalants	Sedatives Main	None
Methamphetamine	None	None

¹ Only the sections listed were flagged by the data diagnostics programs to be examined further. However, once one of these sections was flagged for a respondent, all other non-substance use sections were examined manually for patterns.

² Responses to the back-end demographics, education, and employment (for respondents aged 15 years or older) sections were examined together to trigger a single flag.

2.3.2.1 Investigation of Response Patterns in Initial Substance Use Sections

A data diagnostics program was developed for substance use questions other than prescription drugs to screen for patterns of responses that might raise questions about the validity of the interview as a whole or the validity of responses in specific sections. Cut points for establishing these thresholds for flagging respondents' data for further examination were originally developed for the 1999 survey based on empirical review of data; these cut points have continued to be used since 2015. Cut points for establishing these thresholds differed, depending on the measure. The data diagnostics program flagged the following patterns for substances other than prescription drugs:

- For hallucinogens and inhalants, data were flagged if respondents reported lifetime use of six or more specific drugs in these sections. This pattern could indicate that respondents were indiscriminately keying data or selecting response options without paying attention to how they were answering.
- For hallucinogens and inhalants, data were flagged if respondents entered alternating "yes/no" responses to questions about lifetime use of specific hallucinogens or inhalants. These alternating answers could be patterned.
- For substances other than tobacco or alcohol, data were flagged if respondents reported using three or more drugs every day or just about every day in the past year or past month (where applicable). For this pattern, one might question either the validity of the answers or the respondent's competence to complete the interview.
- For nicotine through methamphetamine, data were flagged if respondents reported first using two or more substances at age 1 or 2. This pattern could indicate indiscriminate choosing of 1s or 2s, especially given that the age-at-first-use questions followed gate questions where a response of 1 denoted "yes" and a response of 2 denoted "no."
- For nicotine through methamphetamine, data were flagged if respondents consistently chose the same code (either 1 or 2) for all questions they were asked in a section (i.e., after lifetime use has been reported), which would suggest a pattern of indiscriminate answering.²²

These patterns of responses to substance use questions other than prescription drugs were examined on a case-by-case basis to determine whether an interview record should be retained as a final respondent or dropped.

A separate data diagnostics program was developed to screen for patterned responses in the prescription drug data. Cut points to establish thresholds for flagging respondents' prescription drug data for further examination were developed based on empirical review of data from the

²² Since 2001, response categories for certain consistency checks have used 4 = yes/6 = no instead of the format of 1 = yes/2 = no. For example, if a respondent reported first using marijuana at age 1 or 2, the respondent could not use a response of "1" or "2" in question MJCC05 to verify that this age at first use was correct. Use of responses of "4" or "6" was designed to stop respondents if they had been engaged in a pattern of keying responses of "1" or "2."

2015 survey, when the prescription drug sections were last redesigned. The following patterns were flagged:

- For each prescription drug, data were flagged if high numbers of prescription drugs were reported in the screener sections for any use in the past 12 months or in the main prescription drug sections for misuse in the past 12 months.²³
- For each prescription drug, data were flagged if responses of 1 or 2 were selected wherever possible in the screener sections. This might indicate that respondents were confusing responses of 1 to mean “yes” and responses of 2 to mean “no,” based on the meanings of these responses in earlier sections (e.g., a response of 2 in question ME01 for methamphetamine meant “no,” but a response of 2 in question PR01 meant used Lortab in the past 12 months).
- For each prescription drug, data were flagged if respondents reported use in the past 12 months of all prescription drugs in a list for screener items that included three or more prescription drugs.
- For each prescription drug, data were flagged if there were miscellaneous answer patterns in the screener sections (e.g., choosing only a response of 1 in PR01, choosing only 2 in PR02, choosing only 3 in PR03, choosing only 4 in PR04).
- For each prescription drug, data were flagged if only responses of 1 were chosen for questions in the main prescription drug sections (e.g., pain relievers) after respondents had reported misuse in the past 12 months of at least one prescription drug in that category.
- For each prescription drug, data were flagged if respondents reported first misuse of two or more prescription drugs at age 1 or 2, which might indicate indiscriminate keying of 1s or 2s.
- For each prescription drug, data were flagged if there was consistent reporting of the same age at first misuse for four or more prescription drugs in a single category (e.g., pain relievers) that respondents misused in the past 12 months.
- For each prescription drug, data were flagged if all possible responses were chosen for the ways in which respondents misused prescription drugs (e.g., question PRYWAYS for pain relievers) or all reasons were chosen for misusing prescription drugs (e.g., question PRYMOTIV for pain relievers) in the past 12 months.

Approximately 10 interviews in 2023 met the usable interview criteria but were treated as nonrespondent interviews because the responses were of questionable validity, based on one or more of the patterns described above. In addition, fewer than 10 interviews in 2023 were retained as respondent interviews, but the original responses to questions in one or more of the sections for alcohol through methamphetamine were replaced with bad data codes. This process included setting responses to bad data that indicated respondents were lifetime users of a given drug. For example, data for some respondents were set to bad data because the respondents provided answers of 1 or 2 to every question that was asked in a section,

²³ Thresholds for high numbers varied by prescription drug section due to the different number of questions in each section. In the screener sections, thresholds of 10 or more for pain relievers and 5 or more for tranquilizers, stimulants, or sedatives in the past year were used.

beginning with the age-at-first-use question. Data for the lifetime variables for these interviews were set to bad data as part of the edits for the lifetime drug use variables (Section 6.2.1). For example, if an interview was identified that had “bad” data for inhalants, the data for the lifetime use of inhalants corresponding to responses in questions IN01A through IN01L were set to bad data as part of the lifetime drug use edits, and a flag was set to indicate that data subsequently needed to be set to bad data for related variables in the inhalants section of the interview and for variables in subsequent sections that were contingent on respondents being users of inhalants.

The changes to the prescription drug questions for 2015 increased the occurrence of patterned responses to prescription drug data. In particular, when respondents answered gate questions in previous sections of the interview, they were instructed that responses of 1 meant “yes” and responses of 2 meant “no.” When respondents reached the prescription drug screeners, however, responses of 1 or 2 corresponded to specific prescription drugs. In question PR01 in the screener for pain relievers, for example, respondents using a keyboard or other device to key numeric responses would key 1 if they used Vicodin® in the past 12 months and would key 2 if they used Lortab® in that period. Respondents would key an answer of 95 if they never used any of the pain relievers in the list for PR01. Improvements to the questionnaire in 2016²⁴ reduced but did not completely eliminate patterned responses in the prescription drug data.

2.3.2.2 Editing of Non-Substance Use and Other Miscellaneous Data Because of Patterned Responses

In the sections of the interview for special drugs through market information for marijuana, two types of edits were made because of patterned responses in the interview. The first type involved edits because of patterned responses in related variables from corresponding substance use sections (see Section 2.3.2.1). When interviews met the usable interview criteria, but variables from one or more initial substance use sections were assigned the appropriate codes for bad data (Section 2.4.3), codes for bad data also were assigned to the corresponding variables in the subsequent sections. This edit was implemented regardless of whether the corresponding variables in these subsequent sections had patterned responses. For example, if a respondent’s variables in the inhalants section were assigned codes for bad data because of patterned responses in that section, and the respondent was routed to questions pertaining to inhalants in the substance use disorder section, then codes for bad data were assigned to the edited variables for inhalant use disorder.

The second type of edit applied to other occurrences of patterned responses specifically in data for the special drugs section through the coronavirus disease 2019 (COVID-19) section. These edits were implemented for around 100 respondents in 2023. Responses were replaced with bad data codes when patterned responses were detected in the sections for special drugs through COVID-19. For example, answers were replaced with bad data codes if respondents keyed answers of “1” (or multiple-digit responses of “1,” where applicable) to every question

²⁴ For example, verification checks were added to the prescription drug screener sections to ask respondents to verify their answers if they entered three responses in a row of only 1 or three responses in a row of only 2 (e.g., if respondents chose only responses of 2 in questions PR01, PR02, and PR03).

they were asked within a section. In the youth experiences section, adolescents who provided answers of “2” to every question they were asked in the section also had their answers replaced with bad data codes. For respondents to answer all possible items as “2” in the youth experiences section would mean that they provided this response to up to 45 items in a row, including that they had engaged in all “problem” activities (such as getting into fights, carrying a handgun, or selling illegal drugs) exactly “1 or 2 times” in the past year. These respondents who answered “2” wherever possible also would have participated in exactly two school-based activities, two community-based activities, two church- or faith-based activities, and two other kinds of activities in the past year.

2.4 General Issues and Principles for Editing NSDUH Data

Previous sections discussed broader data processing and coding procedures in the data as a whole or in large sections, such as the initial substance use sections of the interview. In contrast, this section focuses on the general principles of data processing and editing across and within interview sections. These steps are taken once the initial processing of interviews have been completed (i.e., the procedures described in Section 2.2), after coded variables have been created (Section 2.3.1), and respondents with patterned responses have been identified (Section 2.3.2).

The following specific issues are addressed in this section:

- editing across sections (Section 2.4.1);
- assignment of relevant “not applicable” codes (Section 2.4.2);
- additional assignment of NSDUH missing data codes (Section 2.4.3);
- assignment of codes to edited “enter all that apply” variables (Section 2.4.4);
- “back-editing” based on “OTHER, Specify” data (Section 2.4.5);
- the “lockout” procedure that prevented respondents, interviewers (for in-person data), and other family members (for questions about health insurance and income) from returning to earlier sections of the interview (Section 2.4.6); and
- general principles for the “flag and impute” procedures for substance use variables other than for prescription drugs (Section 2.4.7).

2.4.1 Editing across Sections

An important principle that was followed in editing the data was that responses from one section (e.g., hallucinogens) generally were not used to edit variables in another section (e.g., inhalants). In particular, data following the initial substance use sections (special drugs through COVID-19) were not used to edit related variables in the initial substance use sections (nicotine through sedatives). For example, if respondents reported in the heroin section that they last used heroin more than 12 months ago, but then reported in the later special drugs section that they last used heroin with a needle more than 30 days ago but within the past 12 months, then the heroin recency variable HERREC from the heroin section was not edited to take into account this more recent indication of heroin use from the later section. Rather, HERREC retained the respondents’ initial response that they last used heroin more than 12 months ago. Consequently, the documentation in NSDUH codebooks for the variables in

sections of the interview that follow the initial substance use sections includes a footnote to indicate that these variables may be inconsistent with data in other sections of the interview.

This principle of not using substance use data in later interview sections to edit data in the initial substance use sections has been important for maintaining consistent data to assess trends in substance use.²⁵ If variables in the initial substance use sections were allowed to be edited based on respondents' answers in the later sections, then key substance use estimates could change across years as later questions or entire interview sections were modified, added, or deleted.

The exception to this principle involved situations in which responses in one or more sections governed whether respondents were asked questions in a later section. For example, the alcohol and drug use treatment section was relevant only for respondents who reported some lifetime use of alcohol or other drugs, excluding tobacco products or nicotine vaping. Respondents who reported in the initial substance use sections that they had never used alcohol, marijuana, cocaine, heroin, hallucinogens, inhalants, methamphetamine, or prescription psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives) were not asked corresponding follow-up questions in the alcohol and drug use treatment section. In such cases, during the editing process, blank values in the alcohol and drug use treatment variables were replaced with codes to indicate that respondents were not asked the follow-up questions in this section because they reported never having used any of the relevant substances.

2.4.2 Assignment of Relevant “Not Applicable” Codes

Because the CAI logic controlled whether respondents were skipped out of some questions based on their answers to previous questions, an important aspect of editing the NSDUH data involved replacing blank codes (98s) in the unedited data with appropriate codes to indicate that the questions had been skipped because they did not apply. The following codes were assigned when respondents were skipped out of a given question and it could be determined unambiguously that the respondent had legitimately been skipped out of the question, based on the answer(s) to one or more previous questions:

- 91 (or 991, or 9991, etc.) = NEVER USED [DRUG(S) OF INTEREST];
- 93 (or 993, or 9993, etc.) = USED [DRUG] BUT NOT IN THE PERIOD OF INTEREST; and
- 99 (or 999, or 9999, etc.) = LEGITIMATE SKIP.

Strictly speaking, codes of 91 and 93 in the CAI data could be considered variants of the more generic legitimate skip code. Their use was designed to provide analysts with more information about the reason that respondents were skipped out of a particular question.

Codes of 91 and 93 were used most often in the initial substance use sections of the interview. For example, a code of 91 (or 991, etc.) in the marijuana section denotes the pattern where

²⁵ Maintaining consistent data to assess trends in substance use does not include situations that could affect the comparability of estimates across years, such as the questionnaire changes in 2015 that resulted in new baselines, methodological changes in 2020 in response to the COVID-19 pandemic, or the continued use of multimode data collection in years after 2020.

respondents were skipped out of all remaining marijuana questions because they answered “no” to the lifetime marijuana question MJ01. Similarly, a code of 93 in the marijuana section denotes situations where respondents were lifetime marijuana users but were definitely not users in the past 30 days, past 12 months, or both.

Codes of 91 and 93 also were used to a limited extent in sections of the interview following the initial substance use sections. These codes were used because the CAI logic took into account respondents’ prior answers to the initial substance use questions to determine whether particular questions applied in later sections of the interview. For example, questions about cocaine use disorder were relevant to respondents who used cocaine in the 12 months prior to the interview. Thus, if respondents last used cocaine more than 12 months prior to the interview, a code of 93 in the cocaine use disorder variables would signify to an analyst why the CAI program skipped the respondents out of these questions. Similarly, a code of 91 in the cocaine use disorder variables would signify to an analyst that respondents were skipped out of these questions because they never used cocaine.

A legitimate skip code of 99 was used most often in sections of the interview following the initial substance use sections. For example, the youth experiences section was intended to be administered only to respondents aged 12 to 17. If respondents were aged 18 or older, a code of 99 was assigned in the editing process to the skipped youth experiences variables. Similarly, if adult respondents reported in the mental health section that there was not a time in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days, then the CAI program skipped the respondents out of the questions about symptoms of psychological distress in the past 12 months. A code of 99 was assigned to the skipped variables in this situation to signify that the adult respondents did not have a worse period of psychological distress within the past 12 months than they did in the past 30 days.

The following codes also were assigned through editing:

- 81 (or 981, or 9981, etc.) = NEVER USED [DRUG(S)] Logically assigned;
- 83 (or 983, or 9983, etc.) = USED [DRUG(S)] BUT NOT IN THE PERIOD OF INTEREST Logically assigned; and
- 89 (or 989, or 9989, etc.) = LEGITIMATE SKIP Logically assigned.

These codes were given values in the 80s to signify that existing values were overwritten during the editing process. For example, if a respondent was somehow routed into the youth experiences section, but that respondent was subsequently classified as being 18 or older (Section 4.2.2), then any answers that the respondent gave in the youth experiences section were overwritten with a code of 89 (or 989, etc.). These codes signify that this adult respondent logically was not eligible to be asked the youth experiences questions. Section 6.3 presents examples of edits where codes of 81 or 83 were assigned to prescription drug variables.

However, the codes described in this section were assigned *only* in situations where there was total certainty that a respondent should have skipped a question. For example, if respondents reported in question MJ01 that they never used marijuana, it was clear that subsequent

questions about marijuana use (e.g., age at first use of marijuana) did not apply. See Section 3.1 for a description of imputation indicators, including a brief discussion of the codes assigned to respondents whose imputed values meant that they legitimately skipped out of the corresponding questions.

The CAI skip logic often treated responses of “don’t know” or “refused” to gate questions as equivalent to a negative response. For example, if respondents were uncertain whether they ever used marijuana (and answered question MJ01 as “don’t know”), the CAI program skipped them out of all remaining questions about marijuana use, as though these respondents had never used it. From the standpoint of respondent burden, this routing makes sense. There is little value in asking respondents how old they were when they first used marijuana if they did not know whether they ever used marijuana. Implicit in the age-at-first-use question is that respondents used marijuana at least once in their lives.

Although the CAI program skipped respondents out of questions if they answered a gate question as “don’t know” or “refused” (or gave similar answers on follow-up if they initially refused to answer a gate question), these types of responses to a gate question do not provide analysts with conclusive information about the behavior or outcome of interest. Consequently, such responses could be thought of as potentially affirmative responses, as opposed to inferring that they are negative responses. Moreover, automatically inferring in an analysis that a response of “don’t know” or “refused” is equivalent to a response of “no” will introduce a negative bias in estimates.

As noted previously, respondents who initially refused to answer a question about their lifetime use or nonuse of a drug were given a second opportunity to answer the question as “yes” or “no.” Similarly, if respondents who initially did not know whether they ever used a drug had thought about the issue further, they may have recalled a time when they in fact had used it—and more detailed questions about use of the drug would have been relevant for these respondents. Alternatively, if respondents gave more thought to the issue and decided that they really should answer the lifetime drug use question as “no,” an analyst would have a solid basis for determining that subsequent questions did not apply.

Further, the procedures for statistically imputing missing data did not automatically infer lifetime nonuse when respondents did not provide definite information about whether they had ever used a given drug. Rather, such respondents were eligible to be statistically imputed to be lifetime users or nonusers. For these reasons, variables retained missing values in the editing procedures when questions were skipped due to respondents answering a lead question as “don’t know” or “refused” (or answering in a similar manner in response to a follow-up probe). Stated another way, retaining missing data when respondents answered a lead question for lifetime use of a substance as “don’t know” or “refused” did not result in making the default assumption that these respondents never used the substance; making an inherent assumption of nonuse would introduce a negative bias into the data.

2.4.3 Additional Assignment of NSDUH Missing Data Codes as Part of Editing within a Section

Previous sections described the procedures for assigning missing data codes in the interview data as a whole or in large sections, such as when the interview date was questionable (Section 2.3.3). This section discusses additional principles for assigning (or retaining) missing data codes as part of the editing procedures for a given section.

If respondents refused a gate question or questions that governed a skip pattern in a section, refusal codes were sometimes assigned to all of the subsequently skipped items in the section as part of the editing procedures. In other words, the lead refusal was propagated, and it was logically inferred that the respondents were globally refusing to answer any questions on that topic. Propagation of refusals also has been included less commonly in editing procedures for new questions introduced in the survey after 2015. Codes for “refused” and “blank” both refer to missing data, regardless of which value is assigned. Therefore, respondents are treated the same in subsequent data processing steps (e.g., imputation) for both of these missing values. Propagation of refusals also can be difficult to implement correctly for complex skip logic, with limited benefits to analysts. Propagation of responses of “don’t know” also did not occur when respondents answered a gate question or questions as “don’t know.”

In addition, data sometimes were identified that were inconsistent with other data in a respondent’s record. For the sections of the interview for nicotine through methamphetamine, for example, if respondents reported first using a drug at an age greater than their current age, the CAI program indicated to the respondents that this age at first use was inconsistent. The respondents were prompted to revise the age at first use, their current age, or both, to make the data consistent.²⁶ Respondents did not always resolve these types of inconsistencies. If the age at first use was still inconsistent with the respondent’s age, despite the opportunity that the respondent had to resolve the inconsistency, then a bad data code was assigned to the age-at-first-use variable to indicate that the data were inconsistent with other data. The following codes were assigned to denote “bad data”: 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned. Other situations where bad data codes were assigned are discussed below in connection with specific steps in the editing process.

2.4.4 Assignment of Codes to Edited “Enter All That Apply” Variables

As noted in Section 2.2.2, the initial creation of individual unedited variables for “enter all that apply” questions involved assignment of the relevant code that was shown to respondents during the interview for that question, or a code of 98 (blank) if respondents did not choose that item from the list. This section describes the creation of the edited “enter all the apply” variables. Most edited variables corresponding to “enter all that apply” questions that originated prior to 2013 were coded as 1 if respondents chose that item from the list, and values of 98

²⁶ Since 2015, respondents who reported misuse of a specific prescription drug at an age that was later than their current age have been able to revise their age at first misuse but are not allowed to change their current age.

were recoded with a value of 6. Documentation for these edited values for “enter all that apply” variables was as follows:

- 1 = Response entered, and
- 6 = Response not entered.

For example, underage respondents aged 12 to 20 who drank alcohol in the past 30 days were asked in question CA02B where they were the last time they drank. These respondents could report that they were in one or more of eight places (including “some other place”) when they drank alcohol the last time. The code of 4 that was assigned in the initial processing to the unedited variable from CA02B for drinking at a park, on a beach, or in a parking lot was reassigned a code of 1 as part of the editing procedures. If the respondent did not choose this place from the list in question CA02B but reported another place from the list, then the code of 98 in the variable for drinking at a park, on a beach, or in a parking lot (which had been assigned during the remapping of “enter all that apply” variables) was replaced with a code of 6 during the editing process.

Use of the code of 6 was intended to indicate to analysts that not choosing an available response from the list was not exactly the same as an answer of “no” in questions that required respondents explicitly to answer “yes” or “no” about a behavior of interest. In other words, a response of “no” in a “yes/no” type of question can be thought of as an active indication that the behavior or characteristic of interest did not apply, whereas not choosing a response in an “enter all that apply” list can be thought of as a passive indication that a particular behavior did not apply. In practice, however, not choosing a response from a list often was treated as being equivalent to a response of “no” in subsequent editing steps.

However, this procedure was modified for “enter all that apply” questions that were added to the interview beginning in 2013. These edited variables were coded as 1 if respondents chose that item from the list, and values of 98 were recoded with a value of 2. Documentation for these values was as follows:

- 1 = Yes, and
- 2 = No (not entered).

As for the code of 6 that was described previously, the documentation of “No (not entered)” for the code of 2 was intended to indicate to analysts that not choosing an available response from the list was not exactly the same as an answer of “no” in questions that required respondents explicitly to answer “yes” or “no.”

[Table 2.2](#) illustrates how “enter all that apply” questions were coded for question PR03 in the pain relievers screener. For example, if a respondent reported using generic oxycodone in the past 12 months by selecting response option 2 in question PR03, then the code of 2 that was assigned to the variable for oxycodone use as part of the initial processing was reassigned a code of 1 (see patterns 1 and 3 in [Table 2.2](#)). If the respondent did not choose generic oxycodone from the list of drugs in PR03 but reported using Roxicodone®, then the code of 98 for generic oxycodone was replaced with the code of 2 (see pattern 2 in [Table 2.2](#)). Editing of

the variables for any use of specific prescription drugs in the past 12 months when respondents chose the response of 95 (i.e., meaning that they had not used any of the drugs listed in a given question, as shown in pattern 4 in [Table 2.2](#)) is discussed in Section 6.3.2.

Table 2.2 Response Patterns for Pain Reliever Screener Question PR03

Response Pattern	Raw (i.e., Unedited) Value after Initial Processing ¹			Value after Editing ²	
	Roxicodone® (PR031 - Level 1)	Generic Oxycodone (PR032 - Level 2)	Not Used Either (PR033 - Level 95)	Roxicodone® (ROXIANYR)	Generic Oxycodone (OXYDANYR)
1	1	2	98	1	1
2	1	98	98	1	2
3	98	2	98	2	1
4	98	98	95	95	95
5	94	94	94	94	94
6	97	97	97	97	97

¹ Raw values: 1 = Response entered for Roxicodone; 2 = Response entered for generic oxycodone; 95 = Respondent did not use either drug; 94 = DON'T KNOW; 97 = REFUSED; 98 = BLANK.

² Edited values: 1 = Yes; 2 = No (not entered); 95 = Respondent did not use either drug; 94 = DON'T KNOW; 97 = REFUSED.

2.4.5 “Back-Editing” Based on “OTHER, Specify” Data

“Back-editing” refers to situations in which answers to a given question can be used to make inferences about how another question (often a preceding question) within the same section should have been answered. Specifically, the principles and procedures that are discussed in this section refer to use of “OTHER, Specify” data to go “back” and edit an earlier variable according to what respondents specified in a later series of questions for that section.

A type of inconsistency that could commonly occur in the NSDUH data involved situations in which respondents did not answer a question affirmatively (e.g., in question LS01A, whether they ever used LSD). In the same section, however, they could report something in “OTHER, Specify” items that indicated that question LS01A should have been answered as “yes” (e.g., specifying use of LSD as some “other” hallucinogen that the respondent had ever used). When respondents specified something that corresponded to an item they had been asked about previously, but they had not answered that previous item as “yes,” then the editing procedures assigned a value of “yes” to the relevant question. The following code typically was used when a response of “yes” was logically inferred: 3 = Yes LOGICALLY ASSIGNED.

If there was a lead to the “OTHER, Specify” question that was in the form of a “yes/no” question (e.g., “Have you ever, even once, used any other hallucinogens besides the ones that have been listed?”), the affirmative answer was retained in the lead to the “OTHER, Specify” question (i.e., having ever used any other hallucinogens), in addition to the inference being made that the answer to another question logically was “yes” (e.g., “Have you ever, even once,

used LSD, also called 'acid'?"). The redundant specify code also was retained in the "OTHER, Specify" variable(s) to indicate to analysts the source of the logically inferred "yes" value.²⁷

This principle also applied to the editing of variables for the use and misuse of specific prescription drugs in the past 12 months based on answers in "OTHER, Specify" data for the misuse of other prescription drugs in that period. The lead to the "OTHER, Specify" questions for the misuse of other prescription drugs also were "yes/no" questions (e.g., "In the past 12 months, did you use any other/any prescription pain reliever in a way a doctor did not direct you to use it?"). Therefore, affirmative answers were retained in the lead to the "OTHER, Specify" questions for the misuse of other prescription drugs, in addition to logical inferences being made that respondents should have previously reported the use or misuse of specific prescription drugs in earlier questions.²⁸

Editing of the prescription drug data based on the "OTHER, Specify" data varied according to whether respondents had reported any use of a given prescription drug in the past 12 months. For example, suppose a respondent reported using the pain reliever Norco® in the past 12 months. In this situation, the edited variable NRCOANYR had already been assigned a code of 1 according to the procedures described in Section 2.4.4. If this respondent did not report misusing Norco® in the past 12 months in question PRY03, but specified Norco® as some "other" pain reliever that the respondent misused in that period, then the edited variable for the past year misuse of Norco (NRCONMYR) was assigned a code of 3 to indicate that question PRY03 logically should have been answered as yes.

However, suppose respondents did not choose Norco® from the list of pain relievers in question PR01. In this situation, question PRY03 would have been skipped. If some of these respondents specified Norco® as some other pain reliever that they misused in that period, then it could logically be inferred that they misused Norco® in the past 12 months and also that they used this prescription drug in the past 12 months for any reason. A code of 5 was assigned to the variables for any use and misuse (e.g., NRCOANYR and NRCONMYR for Norco®), where 5 = Yes LOGICALLY ASSIGNED. This code of 5 distinguished the respondents who specified the misuse of a specific prescription drug in the past 12 months but did not report the use of that drug in the past 12 months from other respondents who reported use but not misuse in that period. In this latter situation, the individual edited variable corresponding to the past year misuse of Norco® was assigned a code of 3.

²⁷ Depending on the analysis, therefore, some further handling of data could be warranted to prevent "double counting." For example, if a respondent was logically inferred to be a lifetime LSD user and reported only LSD in the "OTHER, Specify" data for the lifetime use of other hallucinogens, then an estimate for lifetime use of other hallucinogens would be redundant with an estimate for lifetime use of LSD.

²⁸ As for the hallucinogens example, some further handling of data in other analyses could be warranted to prevent double counting. For example, if a respondent was logically inferred to have misused Norco® (a hydrocodone product) in the past year and reported only Norco® in the "OTHER, Specify" data for the misuse of other pain relievers in the past year, then an estimate for the past year misuse of other pain relievers would be redundant with an estimate for the past year misuse of hydrocodone products.

For “enter all that apply” variables other than the screener variables for prescription drugs, a code of 3 typically was used when a response was logically inferred based on “OTHER, Specify” data. For example, if underage respondents aged 12 to 20 drank alcohol in the past 30 days and specified that they drank alcohol at a park, on a beach, or in a parking lot as some other place they drank alcohol the last time they drank, but they did not report drinking alcohol in this place in question CA02B, then a code of 3 was assigned to the variable CADRPUBL from CA02B.

2.4.6 “Lockout” Following Completion of Relevant Sections

Once the self-administered portion of in-person interviews were completed, interviewers and respondents were locked out of this section to provide additional protection of the privacy of respondents’ answers. However, based on answers that respondents gave in the interviewer-administered questions following the self-administered section, interviewers could go back and change information in demographics questions that were in the interviewer-administered section that preceded the self-administered questions. For example, interviewers could discover when they were completing the household roster section of the interview that a female respondent’s gender had been incorrectly keyed as male, or vice versa (Section 4.2.4). The interviewer was allowed to change the information about the respondent’s gender at the beginning of the interview.

If interviewers went back and changed answers to the interviewer-administered demographic questions, however, the CAI program did not reroute respondents back into the self-administered section to answer questions that previously had been skipped. For example, respondents who were reported to be male when they were routed to the ACASI questions on pregnancy were not asked these questions. If interviewers went back and changed respondents’ gender to be female after they had exited the ACASI portion of the interview, then these respondents were not routed back to the pregnancy questions. In this situation, the pregnancy data for respondents whose final gender was reported to be female were assigned codes of 98 (blank) to indicate that data were missing for these questions. If questions were asked but should not have been asked due to a change to the final gender, and the answers remained in the completed data, then the values in the corresponding edited variables were overwritten with codes of 89 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2).

Respondents who completed the web-based interview did not encounter this lockout feature until after they completed the household roster portion of the interview. For web-based interviews, there were no concerns with an interviewer seeing a respondent’s previous answers. However, a family member other than the respondent (i.e., a “proxy”) was allowed to answer the questions about the respondent’s health insurance coverage and the kinds of income the respondent and other family members received.²⁹ Therefore, to protect the privacy of web respondents’ answers, web respondents and other family members who might be answering the

²⁹ For in-person and web interviews, the assumption for questions about health insurance coverage and income was that a family member other than the respondent might be better able than the respondent to provide accurate information about these topics.

remaining questions were not allowed to see previous interview questions once web respondents completed the household roster section.

2.4.7 “Flag and Impute” Principles for Substance Use Data Other than Prescription Drugs

The editing procedures for establishing when respondents last used a substance are critical for creating final published estimates from NSDUH of the prevalence of substance use in the United States. In addition, data from the initial substance use sections on the most recent use are important for establishing whether skipped questions in later sections truly were not applicable or if there might be some question about whether these skipped questions might have applied to the respondent. For example, respondents who reported that they used cocaine in their lifetime but that they last used it more than 12 months ago were not asked questions about cocaine use disorder. However, if any of these respondents also reported that they first used cocaine at their current age, these reports of recent initiation would suggest that they may have used cocaine in the past 12 months, in which case they should have answered the substance use disorder questions for cocaine.

Four possible ways of editing the data for most recent use of a substance were examined as part of the methodological research for data processing and estimation procedures using the CAI data (Kroutil & Myers, 2002). The flag and impute rule for editing the CAI recency-of-use variables for substances other than prescription drugs flags inconsistencies between a recency variable and related variables but does not make a decision about the final recency category. Rather, this rule allows these inconsistencies for recency-of-use and related data to be resolved through statistical imputation.

For example, if respondents originally reported that they last used a substance more than 12 months ago in the nicotine through methamphetamine sections of the interview, but they also reported first using it at their current age, this procedure inferred that the respondents were *at least* lifetime users. In the imputation procedures, these respondents’ data for most recent use were imputed to indicate most recent use in any period (i.e., in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago). Also, the data on past year initiation for substances other than prescription drugs that were inconsistent with most recent use of a given substance more than 12 months ago were set to bad data (Section 2.4.3), which treated the initiation data as missing. Consequently, if respondents were imputed to have last used the substance at some point in the past 12 months (including use in the past 30 days), then they could be imputed to have initiated use of the substance within the past 12 months or more than 12 months ago based on data from corresponding donors whose reports of first use and most recent use were consistent. However, if respondents were imputed to have last used the substance more than 12 months ago, then initiation data from a donor respondent also would be consistent with initiation more than 12 months prior to the interview date.

Therefore, this edit rule does not allow inconsistent data to remain between reports of the most recent use of a substance and other reports of use in that section of the interview. However, this rule does not automatically infer that the respondents last used a substance more recently

than they originally reported, nor does it automatically discount reports of more recent use. Rather, if respondents give an answer within a substance's section that conflicts with their original responses to the recency-of-use question, then the recency variable is statistically imputed using data from a suitable donor record without these inconsistent data (see Chapter 3). If final assignment occurred to indicate use in a more recent period than respondents originally reported, this decision typically was made through statistical imputation rather than deterministic editing.³⁰

The changes to the prescription drug questions beginning in 2015 greatly reduced the opportunity for respondents to give inconsistent answers to these questions. For example, respondents were asked questions about their first misuse of prescription drugs only for the specific prescription drugs that they misused in the past 12 months. Consequently, beginning in 2015, respondents could not report that they initiated the misuse of prescription drugs in a given psychotherapeutic category (e.g., pain relievers) in the past 12 months without also having reported misuse of some prescription drugs in that period. Because the changes to the prescription drug questions reduced the opportunity for respondents to answer the questions inconsistently, these "flag and impute" procedures did not need to be applied to the editing of the prescription drug data. Rather, new "edit rules" were developed for the prescription drug data beginning in 2015 that handled remaining inconsistent patterns in these data that SAMHSA decided would be unallowable in the final analytic dataset (see Section 6.3.13).

³⁰ Limited exceptions that involved deterministic editing of recency variables are discussed in Section 6.2.2.5.

3. Imputation Procedures Used in NSDUH

3.1 Introduction

As with most large-scale sample surveys, the respondent datasets for the 2023 National Survey on Drug Use and Health (NSDUH) contained missing responses for some items, inconsistent or invalid responses, and violations of skip patterns. Although the survey instrument was designed to enforce skip patterns and to perform consistency checks during data collection, invalid and inconsistent responses still occurred. These response errors are a source of bias in the analysis of NSDUH data (Cox & Cohen, 1985).

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

"Imputation" is the term used to describe the replacement of missing data with plausible values. Most commonly, imputation is used when a respondent answers some questions on a survey but not others. This is a condition known as "item nonresponse." By contrast, when a selected individual does not respond to any question on the survey at all, or does not respond to enough key questions for the interview to be useful for research purposes, this is a condition referred to as "unit nonresponse." In such cases, weighting adjustments are normally employed to account for these missing data. As an initial step, prior to any processing of the data, unit nonrespondents were discarded, and only unit respondents (i.e., item respondents and item nonrespondents for any given questionnaire item) were included in the subsequent editing, imputation, and analysis of NSDUH data. Throughout the remainder of this report, the terms "nonresponse" and "nonrespondent" are used to denote item nonresponse and item nonrespondent, respectively.

For the most part, variables that underwent imputation had low item nonresponse rates. In the 2023 NSDUH, 359 variables underwent imputation. Of these, 245 required 3 percent or fewer of their records to be imputed. [Table 3.1](#) is a summary of the weighted percentages of cases imputed by variable group.

Table 3.1 Weighted Percentages of Cases Imputed by Variable Group

Variable Group	Number of Variables	Mean	Minimum	25 th Percentile	Median	75 th Percentile	Maximum
Substance Use Variables ¹	163	2.21%	0.01%	0.49%	1.07%	2.63%	27.71%
Date-of-First-Use Variables	60	87.88%	9.43%	96.74%	98.47%	100.0%	100.0%
Substance Use Treatment Variables	19	1.48%	1.31%	1.42%	1.47%	1.48%	1.88%
Mental Health Variables	37	2.77%	0.82%	2.26%	2.71%	2.87%	5.28%
Mental Health Treatment Variables	17	1.79%	1.69%	1.77%	1.81%	1.83%	1.86%
Demographic Variables	18	1.57%	0.08%	1.06%	1.07%	2.38%	4.10%
Income Variables	10	4.57%	2.82%	3.13%	3.64%	5.64%	10.38%
Health Insurance Variables	6	3.09%	2.57%	2.69%	2.86%	3.07%	4.51%
Roster Variables	6	2.29%	2.14%	2.21%	2.31%	2.37%	2.38%
Pair Variables	23	1.21%	0.13%	0.49%	1.10%	1.78%	2.95%
All except Date-of-First-Use Variables	299	2.19%	0.01%	0.83%	1.67%	2.72%	27.71%
All Variables	359	16.51%	0.01%	1.03%	1.94%	4.53%	100.0%

¹ Substance use variables do not include day-of-first-use, month-of-first-use, or year-of-first-use variables. However, they do include the substance use disorder variables.

Many of the variables with higher imputation rates are those that are subject to more logical constraints and consistency checks than other variables (e.g., frequencies of use or variables related to the age at which a respondent first [mis]used a particular substance). Questions that are often considered to be sensitive, such as Total Family Income (Finer Categories), also have higher imputation rates (10.38 percent in 2023). Questionnaire items that few respondents saw due to negative responses to gate questions, such as substance use disorder questions for rarely used substances, also had higher imputation rates.

The date-of-first-use variables (i.e., year-, month-, and day-of-first-use [or misuse for prescription drugs]) had much higher imputation rates than all other measures that undergo imputation in NSDUH. These higher rates are due to the analytic purpose of these variables and the way in which the necessary information is obtained in the questionnaire. These variables are used to determine past year initiation status (also called “recent initiation”) for each substance rather than to obtain an exact date at which a respondent first used or misused a substance.

For substances other than prescription drugs, respondents who report lifetime use of a substance are asked to provide their age at first use (AFU). If the AFU is within 1 year of the respondents’ current age, the respondents are then asked to provide their year of first use (YFU) and month of

first use (MFU). For all other lifetime users of a substance, the respondents are known to not be recent initiates, so the computer-assisted interviewing logic skips the YFU and MFU questions and these variables are imputed, which leads to high imputation rates. Regardless of the AFU reported by a lifetime user, the questionnaire does not include a question measuring the day of first use (DFU) for any substance. Consequently, the DFU variable is imputed for all lifetime users of a substance, excluding prescription drugs.

For prescription drugs including pain relievers, tranquilizers, stimulants, and sedatives, respondents are asked to report their age at first misuse (AFM) for each specific prescription drug that they misused in the past 12 months. If the respondents could be recent initiates of misuse based on their AFM, the respondents are asked their year of first misuse (YFM) and month of first misuse (MFM). If recent initiation for the overall category based on the AFM, YFM, and MFM data for each specific prescription drug misused in the past 12 months still cannot be determined, the respondents are specifically asked whether any of their misuse occurred more than 12 months prior to the interview date. The date-of-first-use variables are then imputed for the overall category (e.g., pain relievers) for recent initiates of misuse only, rather than for all lifetime users. As with the nonprescription drugs, the day of first misuse is never asked of any respondents. These variables are created solely for determining recent initiation of misuse and have imputation rates of 100 percent.

Despite the high imputation rates for the date-of-first-use variables, the primary outcome of interest—recent initiation status—does not require imputation for most people in the domain. For most lifetime users of a substance, the initiation status can be determined based on other available information (e.g., recency of use, AFU/AFM) provided by the respondent.

Once processed, imputed values cannot be distinguished from nonmissing values for a given variable in the final dataset. Therefore, observations with imputed data must be identified with a concomitant indicator variable. The majority of imputation-revised variables for the 2023 NSDUH have the prefix “IR” attached to their names.³¹ Each imputed variable has an associated indicator variable, identified by the prefix “II” or “II2,” that can be used to identify which values were imputed and which were not.³² The levels of a typical imputation indicator are as follows:

- 1 = From questionnaire
- 2 = Logically assigned
- 3 = Statistically imputed
- 9 = Legitimate skip

Assignment of a code of 9 to an imputation indicator reflected the prior assignment of a legitimate skip code as part of the editing process (Section 2.4.2). Also, if a question had

³¹ Although no missing data were possible for gender because a response to this item was required before the interview could proceed, the “IR” prefix for IRSEX was maintained for continuity with past years.

³² The indicator with the prefix “II2” provides more details about the source of the imputed or logically assigned value. However, for many imputed variables, this additional information is contained in the indicator with the prefix “II.”

missing or inexact data and this question governed the skip logic for subsequent questions, respondents typically were skipped out of the subsequent questions. Respondents requiring imputation for the variables that governed a skip pattern typically receive a value of 3 (statistically imputed) for the imputation indicators associated with both the governing variables and the imputed variables that were nested within the skip pattern. For example, if respondents had missing data for whether they had ever used a particular substance and were imputed to have never used it, the imputation indicators for recency of use of the substance, AFU, frequency of use in the past 12 months (if applicable), and frequency of use in the past 30 days (if applicable) all were assigned a code of 3.

Various methods of imputation have been used since NSDUH was first administered in the early 1970s. With the expansion of the NSDUH sample size in 1999, the predictive mean neighborhood (PMN) method for imputation was implemented. In 2015, the modified predictive mean neighborhood (modPMN) imputation methodology was implemented. Both PMN and modPMN are designed to incorporate the complex interrelationships among items in the current NSDUH, thus maintaining data consistency within individual respondent records.

With any method of imputation, missing responses for a particular variable (hereafter, termed “base” variable) are replaced by values from similar respondents with respect to a number of characteristics (hereafter, “auxiliary variables”). If “similarity” is defined in terms of a single predicted value from a model, these auxiliary variables can be represented by that value. The respondent with the missing value for the base variable is called the “recipient,” and the respondent from whom values are borrowed to replace the recipient’s missing value is called the “donor.” Donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (i.e., the donor has complete data for that variable, and the recipient does not). The term “hot deck” is used to refer to imputations made on recipient base variables using donor values from the same dataset. The PMN methodology, and modPMN methodology for continuous variables, utilized in NSDUH are specialized hot-deck methods and are described in greater detail later in this chapter. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

For the 2023 NSDUH, the only imputations that did not incorporate either the PMN or modPMN methodologies were those used for the birth date, date of first (mis)use,³³ and AFM for prescription drugs. Imputation procedures for the birth date are described in Section 4.2.5. Methods used to assign the date of first (mis)use and AFM for prescription drugs are described in Section 6.5. Two other hot-deck methods—unweighted sequential hot deck (USHD) and weighted sequential hot deck (WSHD) (Cox, 1980, pp. 721-725; Iannacchione, 1982)—were

³³ The date of first (mis)use variables are randomly assigned based on the imputation-revised age-at-first-use variable for nonprescription drugs or the imputation-revised past year initiation status variable for prescription drugs. Additional information about the interview date, birth date, imputation-revised recency and frequency of (mis)use variables, and month and year of first (mis)use is also used to define upper and lower bounds on the date of first (mis)use. Once the earliest and latest possible dates are determined, a day is randomly selected from this interval.

used in past surveys.³⁴ See Chapter 3 of the *2019 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 10: Editing and Imputation Report* (Center for Behavioral Health Statistics and Quality, 2021) for a discussion of the features and limitations of USHD, WSHD, and the random nearest neighbor hot deck (NNHD) methodologies.

3.2 Development of the Predictive Mean Neighborhood Methodology

The PMN methodology developed for and implemented in the 1999 NSDUH was an attempt to address the shortcomings, while retaining the positive characteristics, of the hot-deck imputation methods used in NSDUH prior to 1999. PMN is currently used for variables related to demographic characteristics of the respondent, personal and family income, health insurance coverage, and household composition. The PMN method is also used to impute measures of substance use for alcohol, tobacco products (i.e., cigarettes, cigars, pipes, and smokeless tobacco), marijuana, marijuana ways of use, cocaine and crack cocaine, and heroin.

The PMN method combines two commonly used imputation methods: non-model-based NNHD (Little & Rubin, 1987, p. 65) and a modification of Rubin's model-assisted predictive mean matching (PMM) method (Rubin, 1986), which matches a missing value to the observed value with the closest predicted mean. The PMN method enhances Rubin's PMM method, in that PMN can be applied to both discrete and continuous variables, either individually or jointly. PMN also enhances the NNHD method for discrete variables so that the distance function used to find neighbors is no longer in terms of the original predictor variables and therefore does not require arbitrary scaling.

In addition, the PMN methodology offers the following advantages:

- **A large number of auxiliary variables may be used to determine donors.** Using a model-based hot-deck technique like PMN allows auxiliary variables to be incorporated in two ways: first, as covariates in models, and second, in likeness constraints³⁵ applied to potential donors. Under PMN, the donors need only be "close" to the recipients with respect to the predicted values determined by the models, even when the models include numerous covariates. Moreover, PMN ensures that a sufficient number of potential donors comprise the donor neighborhood so that likeness constraints may be applied on the donor set as needed.
- **Relative importance of auxiliary variables is determined by standard estimating equation techniques.** In PMN, objective criteria based on a rigorous methodology (i.e., regression) quantify the relationship between a given covariate and

³⁴ The USHD method was used exclusively for the 1991-1998 surveys, the paper-and-pencil interviewing sample from the 1999 survey, and all demographic variables in the computer-assisted interviewing sample from the 1999 survey. In the 2002-2003 surveys, missing values in the immigrant variables required WSHD imputation. Note, however, that the USHD and WSHD methods have not been used for NSDUH since the 2000 and 2003 survey years, respectively.

³⁵ Likeness constraints are flexible constraints that govern the similarity between donors and recipients. See Section 3.3.3.1 for details.

the response variable in the presence of other covariates so that the response variable itself is indirectly used to determine donors. Further, the analysis weights can be incorporated in PMN regression models without difficulty.

- **Internal consistency of the post-imputation record is guaranteed.** In PMN, the donor pool can be restricted to those making the post-imputation record logically consistent. For example, if a recipient must receive a cocaine past year frequency of use between 30 and 50, the donors can be restricted to ensure that the recipient receives such a value.
- **Correlations across response variables are accounted for by making the imputation multivariate.** Discrete and continuous variables can be handled jointly and relatively easily in PMN by using the idea of sequential univariate modeling. Further, differential weighting factors can be objectively assigned to different elements of the predictive mean vector depending on the variability of predicted means in the dataset.

3.3 Implementation of the Predictive Mean Neighborhood Methodology

The implementation of PMN on the NSDUH involves three basic steps: response propensity (RP) adjustment, prediction (PRD) modeling, and hot-deck imputation. At the most basic level, the RP adjustment reallocates the analysis weights of the item nonrespondents to item respondents; the PRD model calculates predicted means for both; and the hot-deck step assigns final values to the item nonrespondents based on a distance function derived from these predicted means. These steps are described in more detail in the following sections and are combined in three different ways, called PMN “types” (Section 3.4), to complete the imputation procedures.

3.3.1 Step 1: Response Propensity Adjustment

Response propensity is defined as the probability of response, whether at the unit level or item level. The purpose of response propensity is to adjust the analysis weights for item nonresponse so that the item respondent weights that are used only during the imputation process are representative of the entire domain of interest. In the RP step of PMN, the item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special case of the generalized exponential model (GEM)³⁶ developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

There are four key inputs to the item RP modeling step:

1. **Analysis weights.** For all imputation procedures, the best available analysis weight is used as an input to the NSDUH imputation procedures. Because of the timing of 12-month processing and, in particular, the coordination between the weighting and

³⁶ The GEM macro, which was written in SAS/IML[®] software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of the 2023 Person-Level Sampling Weight Calibration report (Center for Behavioral Health Statistics and Quality, 2025b).

imputation tasks in each NSDUH year, most variables that undergo imputation utilize the preliminary analysis weight developed by the NSDUH weighting team for imputation purposes. The dwelling unit-level design weights were adjusted for nonresponse using a weighting class method by state. The nonresponse-adjusted dwelling unit weights and the person-level design weights were further adjusted for nonresponse at the person level using a weighting class method by state and age groups (12-17, 18-25, 26-34, 35-49, and 50+). The preliminary analysis weight is the product of all design weights and the two nonresponse adjustment factors. For those variables that are processed later in the annual cycle, the final analysis weight may be used instead if it is available at the time of imputation processing. The pair variables described in Chapter 14 utilize yet another weight developed specifically for responding pairs within the same household. See the *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 11: Person-Level Sampling Weight Calibration* report (Center for Behavioral Health Statistics and Quality, 2025b) for full descriptions of the final analysis weights. See the *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 12: Questionnaire Dwelling Unit-Level and Person Pair-Level Sampling Weight Calibration* report (Center for Behavioral Health Statistics and Quality, 2025c) for full descriptions of the pair weights.

2. **Domain indicator.** In this report, a “domain” is defined as the set of respondents who are included in models and for whom predicted means are calculated. For many NSDUH variables that undergo imputation, the domain includes all unit respondents. For others, the domain is a subset of unit respondents. For example, the marital status question is only asked of respondents aged 15 or older. In this case, the domain indicator is set to zero for respondents aged 12 to 14 and to one for respondents aged 15 or older.
3. **Item response indicator.** The item response indicator is set to zero for item nonrespondents and one for item respondents. GEM uses this indicator to identify item nonrespondents and item respondents when reallocating the weights appropriately.
4. **Covariates.** GEM uses a predetermined list of covariates to allocate weights from item nonrespondents to item respondents. The covariates tend to be variables that are correlated with (1) the propensity to respond, (2) the variable that is undergoing imputation, or (3) both. The goal is to avoid bias in the PRD models by allocating the weights of the item nonrespondents to “similar” item respondents, without too greatly inflating the variance of the estimates that utilize these weights (Center for Behavioral Health Statistics and Quality, 2025b). Appendix D provides information on the covariate lists for RP models fit in NSDUH imputation procedures.

3.3.2 Step 2: Prediction Modeling

Utilizing the RP-adjusted weights that were derived in the previous step, the PRD model calculates predicted means, which are used in the hot-deck step(s) to create neighborhoods and select donors. The dependent variable in the model is usually the variable, or some transformation of that variable, that is undergoing imputation. Each model is built using only those cases within the domain with complete responses for that item. Predicted means are then calculated for all domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, logistic regression models are used for the PRD models. For continuous variables, linear regression models are fit. For count variables, Poisson regression models are used. For response variables that are proportions (e.g., months on welfare; see Section 12.4.1.4), a logit transformation is applied to the proportion, and a linear regression model is utilized. The variable sets in which some transformations of the response variables were implemented include the self-administered demographics (Section 5.3.3.2), drugs (Section 6.5), and income (Section 12.4.1.4).

The goal of the PRD model is to protect against bias for as many analysis domains as possible, so these models tend to start with long lists of covariates (Allison, 2008b).³⁷ Appendix D provides information on the covariate lists for PRD models fit in NSDUH imputation procedures. In contrast to explanatory (association) models where model parsimony is a relevant metric of a model's appropriateness, the focus in a prediction setting is on the predicted values only.

The SUDAAN® software package³⁸ is used to fit nearly all PRD models used in the NSDUH.³⁹ Under PMN, all covariates from the applicable starter list are utilized unless SUDAAN produces warning messages, which indicate nonconvergence or model instability.⁴⁰ Generally, in these cases, covariates whose regression coefficients have high p-values are dropped until SUDAAN no longer produces these warning messages. The primary advantage of using SUDAAN to fit PRD models is that the standard errors associated with the regression coefficients properly account for the complex survey design. The predicted means are the same using SUDAAN as they are using, for example, the analogous base SAS® procedure (given the same set of covariates),⁴¹ but the decision on which covariates to drop in the event of model instability or nonconvergence is more informed under SUDAAN because the standard errors, and corresponding test statistics and p-values for model coefficients, properly account for stratification and clustering.

For some logistic regression models, the warning messages produced by SUDAAN may be triggered when a cross-classification of the outcome variable and a covariate has empty or nearly empty cells. Covariates of this type are highly correlated with the outcome variable but cannot be used in the PRD model. However, they are often used in the hot-deck step to identify suitable donors. For the types of regression-based PRD models used for each variable that underwent imputation using PMN, see [Table 3.2](#).

³⁷ Allison (2008b) writes that "... imputation models should be relatively 'rich' so that they may be congenial with lots of different models that could be of interest."

³⁸ SUDAAN® is a registered trademark of Research Triangle Institute.

³⁹ The only exception is the finer income categories, described in detail in Section 12.4.2.

⁴⁰ For some variables, a more automated approach to model fitting is used, which may result in variables being dropped even in the absence of warning messages. For more information, see Section 3.9.

⁴¹ SAS® software is a registered trademark of SAS Institute Inc.

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood

Variable	Domain	Type of Regression Model	SAS®/SUDAAN® Procedure ¹
Demographics			
Marital Status	15 years or older	Multinomial Logistic	MULTILOG
Race	All respondents	Multinomial Logistic	MULTILOG
Hispanic or Latino Indicator	All respondents	Binomial Logistic	RLOGIST
Hispanic or Latino Group	Hispanics	Multinomial Logistic	MULTILOG
Education Level	All respondents	Multinomial Logistic	MULTILOG
Employment Status	15 years or older	Multinomial Logistic	MULTILOG
Immigrant Status: Born-in-U.S. Indicator	All respondents	Binomial Logistic	RLOGIST
Immigrant Status: Age of Entry	Not born in U.S.	Simple Linear	REGRESS
Substance Use²			
Lifetime Substance Use	All respondents ³	Binomial Logistic	RLOGIST
Recency of Substance Use, "Hierarchical" Substances	All lifetime users for past year vs. not past year; all past year users for past month vs. not past month	Binomial Logistic	RLOGIST
Recency of Substance Use, Pipes	All lifetime users	Binomial Logistic	RLOGIST
Recency of Substance Use, Other Substances	All lifetime users	Multinomial Logistic	MULTILOG
12-Month Frequency of Substance Use	All past year users	Simple Linear	REGRESS
Daily Substance Use over Past 30 Days, Cigarettes	All past month users	Binomial Logistic	RLOGIST
30-Day Frequency of Substance Use, Cigarettes	All past month users except those who used daily over the past 30 days	Simple Linear	REGRESS
30-Day Frequency of Substance Use, Other Substances	All past month users	Simple Linear	REGRESS
Age at First Substance Use	All lifetime users	Simple Linear	REGRESS
Household Composition			
Total Number of Rostered People	All respondents	Poisson	LOGLINK
Total Number of Children Younger Than 18	All respondents	Poisson	LOGLINK
Total Number of People Aged 65 or Older	All respondents	Poisson	LOGLINK

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS®/SUDAAN® Procedure ¹
Indicator of Whether the Respondent Has Family Members in Household	All respondents	Binomial Logistic	RLOGIST
Total Number of Respondent's Family Members in the Household (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Total Number of Respondent's Family Members in the Household Younger Than 18 (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Income			
Source of Income	All respondents	Binomial Logistic	RLOGIST
Months on Welfare	All respondents who received welfare payments or welfare services in the past year	Simple Linear	REGRESS
Total Income (Binary)	All respondents	Binomial Logistic	RLOGIST
Finer Category Income	All respondents	Time-to-Event (Survival)	LIFEREG
Health Insurance			
Specific Types of Health Insurance Coverage	All respondents	Binomial Logistic	RLOGIST
Any Other Health Insurance	All respondents with none of the specific types of health insurance	Binomial Logistic	RLOGIST

¹ See RTI International (2013) for more information on all SAS®-callable SUDAAN® procedures in this table except LIFEREG. See SAS Institute Inc. (2014) for more information on the LIFEREG procedure. SAS software is a registered trademark of SAS Institute Inc. SUDAAN is a registered trademark of Research Triangle Institute.

² The substances imputed using the predictive mean neighborhood methodology include alcohol, cigarettes, cigars, pipes, smokeless tobacco, marijuana, marijuana ways of use, cocaine/crack, heroin, and lifetime use of hallucinogen "child" drugs (i.e., LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, *Salvia divinorum*, and "any other" hallucinogen).

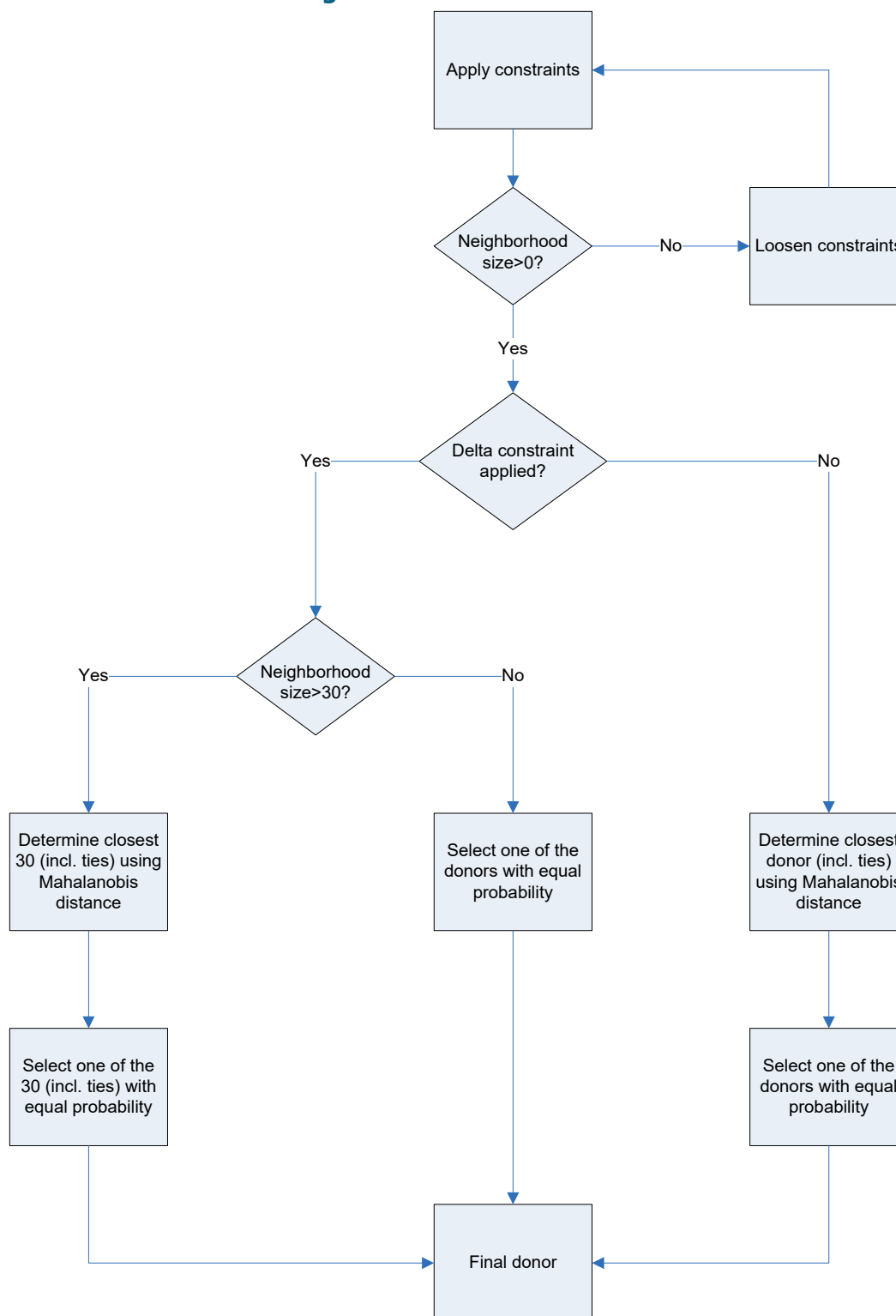
³ The domain for lifetime use of hallucinogen child drugs is defined by the imputation-revised variable for lifetime use of overall hallucinogens.

3.3.3 Step 3: Hot-Deck Imputation

After analysis weights have been appropriately adjusted in the RP step and predicted means have been calculated in the PRD step, the hot-deck step⁴² of PMN is applied to select a donor for each item nonrespondent. The algorithm used to select donors is graphically displayed in the flowchart in [Figure 3.1](#). Briefly, likeness constraints are loosened in an iterative fashion until PMN yields a nonempty donor neighborhood. Mahalanobis distance is then used to rank donors by closeness to the item nonrespondent, and a final donor is selected at random for a given recipient. Many of the hot-deck components used in PMN are described below and appear explicitly in the tables in Appendix E.

⁴² Under PMN, there are two situations on the NSDUH in which the imputation is not a hot-deck step, but is a stochastic imputation based solely on the predicted mean(s) of the recipient, of the type described in Section 5.1 of the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* report (Center for Behavioral Health Statistics and Quality, 2017). These ideas have their origin in Singh et al. (2004), where “Centered PMN” is discussed as an alternative to PMN. These exceptions include the provisional imputation step for imputation set 1 for the health insurance variables (Section 13.3.1.3) and the provisional imputation step for hallucinogen child drugs (Section 6.5.1.2).

Figure 3.1 Donor Selection Algorithm



3.3.3.1 Logical and Likeness Constraints

Logical constraints and likeness constraints are restrictions placed on the set of donors to make imputed values consistent with preexisting, nonmissing values of the item nonrespondents (recipients) and to make candidate donors as much like the recipients as possible. Logical constraints are fixed constraints that prevent logical inconsistencies between variables, and likeness constraints are flexible constraints that govern the similarity between donors and recipients.

The logical constraints are never removed, because to do so would risk the selection of a donor that produces an inconsistent post-imputation record. For example, for the employment status variable, if the item nonrespondent is known to be employed, but full-time vs. part-time status is unknown, the imputed value must come from a donor who is employed as well.

Likeness constraints are placed on the pool of donors to make the attributes of the neighborhood as close as possible to those of the recipient. For example, age and employment status are correlated. A likeness constraint exploits this correlation by requiring the donor's age to be within 5 years of the item nonrespondent's age, but likeness constraints may be loosened if they happen to force the donor pool to be empty.

One likeness constraint that is used in all hot-deck steps, regardless of the variable being imputed, is the delta constraint. This particular likeness constraint requires the donor's predicted mean to be within 5 percent (delta) of the item nonrespondent's predicted mean for each element of the predictive mean vector. If the predicted means are probabilities, the values of delta vary depending upon the value of the predicted mean.

Each delta is defined as 5 percent of the recipient's predicted probability (p) from the PRD model. If the recipient's predicted probability is less than 0.5, the probability is multiplied by 0.05 (i.e., $0.05p$). This value is then used to determine an upper bound (i.e., $p + 0.05p$) and lower bound (i.e., $p - 0.05p$) for the constraint. If the donor's predicted probability is outside these bounds, the donor is eliminated from the donor pool. If the recipient's predicted probability from the PRD model is greater than 0.5, the predicted probability is subtracted from 1 prior to multiplying by 0.05 (i.e., $0.05[1 - p]$). This allows for a looser delta for predicted probabilities close to 0.5 and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is shown in [Table 3.3](#). Logical constraints and likeness constraints, including the order in which likeness constraints are loosened for some variables, are presented in the tables in Appendix E.

Table 3.3 Values of Delta for Various Predicted Probabilities

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

NOTE: In very rare cases, due to transformations of the response variable for the prediction model, p was slightly less than 0 or slightly greater than 1. This occurred only for 30-day frequency and 12-month frequency. If p was less than 0, delta was set to $0.05|p|$. If p was greater than 1, delta was set to $0.05(p - 1)$. See Sections 6.5.1.3.6 and 6.5.1.6.2 for a detailed description of these transformations and their associated back-transformations.

3.3.3.2 Predictive Mean Vector

The predicted means, or expected values from the regression model fit in the PRD step, play a central role in the donor selection algorithm depicted in [Figure 3.1](#). The predicted means are incorporated through the construct of the predictive mean vector. The predictive mean vector is essentially a list of predicted means from the PRD modeling step. In simple cases, the predictive mean vector contains only one element, such as the predicted age at which a respondent began using a drug. In more complex cases, the predictive mean vector includes several elements from several different PRD models, such as the predicted recency and predicted frequency of use for a given drug.

When the PRD model is a logistic regression model, predicted means are calculated for each level of the outcome variable. For example, the employment status variable that undergoes imputation has four levels: employed full time, employed part time, unemployed, or other. Therefore, a single PRD model is fit using a four-level outcome variable, yielding predicted probabilities for each level, as follows:

- E1: P(respondent is employed full time)
- E2: P(respondent is employed part time)
- E3: P(respondent is unemployed)
- E4: P(other)

Note, however, that the predictive mean vector for the employment status variable contains only three elements. It does not include the predicted probability for the “other” level, which in this case is the reference cell implicitly defined by the presence of the other three predicted means.

Occasionally, the predicted means are adjusted so that they are made conditional on what is known for a given respondent. Continuing the example above, some respondents report that they have a job but are unclear about the number of hours they usually work in a week. Because the NSDUH definition of part-time versus full-time employment status was based on working a minimum of 35 hours in a usual week, the predictive mean vector is made conditional on employment of any sort for these respondents. Therefore, the single predicted mean used for these respondents is a conditional probability equal to $E1 \div (E1 + E2)$ or $P(\text{respondent is employed full time} \mid \text{respondent is employed})$. Conditional probabilities are also used in the binary income hot-deck step and the drug recency/frequency hot-deck steps. Predictive mean vectors are presented in Appendix E.

3.3.3.3 Univariate vs. Multivariate Matching and Assignment

If the predictive mean vector consists of only one element, univariate matching is used to select a donor. If the predictive mean vector consists of more than one element, multivariate matching is used to select a donor. The donor may also give values to the item nonrespondent for one variable, a situation known as univariate assignment. Similarly, if the donor provides values for more than one variable, the hot-deck step uses multivariate assignment. [Table 3.4](#) shows examples of NSDUH variables that were imputed using each of the four combinations of univariate/multivariate matching and assignment.

Table 3.4 Examples of Variables Imputed Using Each of the Four Combinations of Univariate/Multivariate Matching and Assignment

Matching Type	Variables Imputed One at a Time (Univariate Assignment)	Variables Imputed in a Set (Multivariate Assignment)
Predictive Mean Vector Has One Element (Univariate Matching)	Hispanic/Latino Origin (Section 4.3.3)	Finer Category Income (Section 12.4.2)
Predictive Mean Vector Has More Than One Element (Multivariate Matching)	Marital Status (Section 4.3.1)	Lifetime Substance Use (Section 6.5.1.1)

Whether the hot-deck step employs univariate or multivariate matching, Mahalanobis distance is used to rank the donors by “closeness” to the item nonrespondent. The Mahalanobis distance is used instead of Euclidean distance in order to standardize the distance in terms of the population variances and covariances of vector components.⁴³ It is given by

$$\sqrt{(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})' \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})},$$

where $\boldsymbol{\mu}_R$ refers to the predictive mean vector for a given item respondent, and $\boldsymbol{\mu}_{NR}$ is the predictive mean vector for a given item nonrespondent. The matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the predictive mean vector, using the set of item respondents that comprise that domain. Because the square of the Mahalanobis distance is a monotone function of the distance itself, and only the ranking of the donors (instead of the absolute distance measure) is used in the algorithm, the additional step of taking the square root of the squared distance is not performed.

3.3.3.4 Missingness Patterns

For many variables imputed on the NSDUH, item nonrespondents were segregated into patterns of nonresponse called missingness patterns. Missingness patterns arise in two ways. First, for sets of variables that underwent multivariate assignment, item nonrespondents were segregated into missingness patterns based on which variables were missing. Second, a new missingness pattern could emerge when logical editing restricted an item nonrespondent to only a subset of the variable’s possible values. The example for employment status discussed above applies here as well: respondents whose employment status was completely unknown had a different missingness pattern than did those who were known to be employed. Often, different predictive mean vectors were used, and different constraints were applied, for different missingness patterns. Many of the tables in Appendix E are segregated by missingness pattern for this reason.

⁴³ When univariate matching is employed, the ranking of donors by Mahalanobis distance is equivalent to the ranking of donors by Euclidean distance. This is not necessarily true when multivariate matching is employed.

3.3.3.5 Final Assignment of Donor Values

Logical and likeness constraints are used to form a neighborhood of potential donors from the pool of item respondents within each missingness pattern. Logical constraints are always imposed to maintain internal consistency, whereas likeness constraints are removed or relaxed in a predetermined order until this donor neighborhood is nonempty. Once a nonempty neighborhood is found, the rest of the PMN donor selection algorithm depends on whether the delta constraint was applied.

If the delta constraint is applied, all members of the neighborhood are similar to the recipient with respect to the predictive mean vector. The final donor is then randomly selected with equal probability from among the “closest” (in terms of Mahalanobis distance) 30 members of the neighborhood. Potential donors whose Mahalanobis distance from the recipient are equal (“ties”) are accounted for in the donor selection algorithm depicted in [Figure 3.1](#). If, on the other hand, the delta constraint is not applied, to ensure that the final donor is as close to the item nonrespondent as possible with respect to the predicted means, the donor with the smallest Mahalanobis distance is selected as the final donor. If there is more than one “closest” donor (i.e., there are ties), the final donor is randomly selected with equal probability from among the closest donors. At the conclusion of the hot-deck step, the item nonrespondent receives values from the selected donor for a single variable (univariate assignment) or for a set of variables (multivariate assignment).

3.4 Predictive Mean Neighborhood “Types”

There are three types of PMN as applied on the NSDUH: Type 1, single RP/single PRD (Section 3.4.1); Type 2, multiple RP/multiple PRD (Section 3.4.2); and Type 3, single RP/multiple PRD (Section 3.4.3). Each of the three PMN types is a coordinated application of the three basic steps of PMN discussed in Section 3.3.

In PMN, an imputation “set” is a set of variables for which a single donor is used in the final hot-deck step.⁴⁴ Sets are formed based on the extent of correlation among variables and the level of missingness in the data. Variables with few missing values and no strong relationships with other variables tend to be processed in an imputation set by themselves. Closely related variables tend to be processed together in the same set to preserve, as much as possible, correlations between variables in the data. However, the more variables that are included in a multivariate set, the less likely it is that a nonempty neighborhood can be found using the delta constraint. Even though there are many advantages to using a multivariate imputation set, one disadvantage in several instances is not being able to apply the delta constraint. [Table 3.5](#) lists the imputation sets for each variable group imputed using the PMN methodology discussed in this report and the PMN type used to process each set.

⁴⁴ Section 3.4.2 defines and discusses the differences between provisional and final hot-deck steps in the context of PMN.

Table 3.5 PMN Types Applied to Each Variable Group and Imputation Set

Variable Group	Imputation Set	PMN Type ¹	Section
Interviewer-Administered Demographics	All (5 sets)	1	4.3
Self-Administered Demographics	All (3)	1	5.3
Substance Use ²	Lifetime (2)	3	6.5.1
	Recency of Pipe Use	1	6.5.1
	Recency/Frequency, Other Substances (7)	2	6.5.1
	Binge Alcohol Use	1	6.5.1
	Marijuana Ways of Use (2)	3	6.5.1
	Cigarette Ever Daily Used	1	6.5.1
	Age at First Use (7)	1	6.5.1
	Age at First Daily Cigarette Use (2)	1	6.5.1
Roster	All (6)	1	11.3.2
Income	Binary	3	12.4.1
	Finer Categories	1	12.4.2
Health Insurance	Types of Health Insurance	3	13.3.1
	Any Other Health Insurance	1	13.3.2
Roster Pair	Pair Relationship	1	14.5.2
	Multiplicities (6)	1	14.5.3
	Household Counts, Sibling-Sibling and Spouse-Spouse (4)	1	14.5.4
	Household Counts, Parent-Child	2	14.5.4

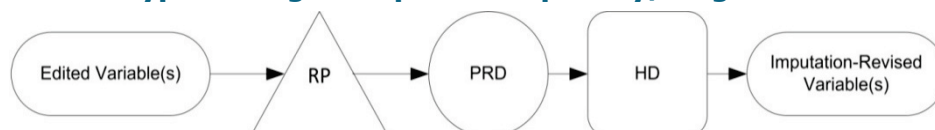
PRD = prediction; PMN = predictive mean neighborhood; RP = response propensity.

¹ PMN Type 1 includes a single RP model and a single PRD model. PMN Type 2 includes multiple RP models and multiple PRD models. PMN Type 3 includes a single RP model and multiple PRD models.

² The substances imputed using the PMN methodology include alcohol, cigarettes, cigars, pipes, smokeless tobacco, marijuana, marijuana ways of use, cocaine/crack, heroin, and lifetime use of hallucinogen “child” drugs (i.e., LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, *Salvia divinorum*, and “any other” hallucinogen).

3.4.1 Type 1: Single Response Propensity/Single Prediction

PMN Type 1, the single RP/single PRD type, involves a single iteration of the three basic steps described in Section 3.3: RP, PRD, and hot-deck imputation. Many variables that undergo imputation using the PMN methodology in the standard processing cycle use this type, including all the demographics and roster variables and the age-at-first-use variables. [Figure 3.2](#) illustrates the single RP/single PRD type of PMN imputation.

Figure 3.2 PMN Type 1: Single Response Propensity/Single Prediction


HD = hot deck; PRD = prediction; RP = response propensity.

Usually the single RP/single PRD type involves univariate assignment in the hot-deck step,⁴⁵ but it may involve univariate or multivariate matching, depending on the PRD model. If the PRD model is a dichotomous logistic regression, linear regression, or Poisson regression model, univariate matching is used because the model produces only one predicted mean. If, on the other hand, the PRD model is a multinomial logistic regression model, multivariate matching is used because the model produces more than one predicted mean (i.e., the predicted probability associated with each level of the response variable). In either implementation, there is only one PRD model.

In the single RP/single PRD type with univariate assignment, the item response indicator is based on the single variable that is being assigned in the hot-deck step. If the single variable is missing, the indicator will denote item nonresponse; otherwise, the indicator will denote item response. In multivariate assignment, the indicator will denote item response if all variables that are assigned in the hot-deck step are nonmissing.

3.4.2 Type 2: Multiple Response Propensity/Multiple Prediction

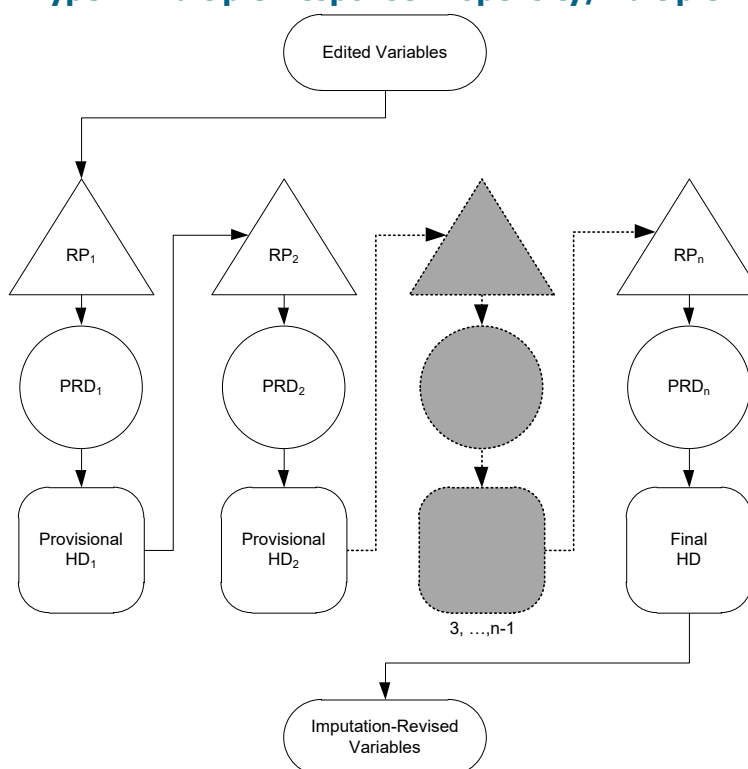
PMN Type 2, multiple RP/multiple PRD, involves multiple iterations of the single RP/single PRD type. However, for all iterations except the last, the hot-deck step is provisional instead of final and involves univariate matching and univariate assignment.⁴⁶ These provisional hot-deck steps tend to be straightforward with respect to constraints and predictive mean vectors, because their only purpose is to fill in missing values so that variables earlier in the sequence can be used as covariates in the RP and PRD models for variables later in the sequence.⁴⁷ In the last iteration, a final hot-deck step is completed, where final imputed values are assigned for all variables involved in the models. This final hot-deck step usually involves multivariate matching and multivariate assignment, depending on which variables within the imputation set are missing. The predicted means from all PRD models are used in this final hot-deck step, and a single record is used to fill all the missing values, thus preserving the relationships among the variables of interest. This is the most refined type of PMN. The recency and frequency variables for a given substance that are imputed using the PMN methodology follow this type. [Figure 3.3](#) illustrates the multiple RP/multiple PRD type of PMN imputation.

⁴⁵ Finer income categories is an example of an imputation set that uses the single RP/single PRD type, but its hot-deck step utilizes multivariate assignment. If the item nonrespondent is missing the finer income category at both the personal and family level, the donor will provide values for both variables in a single hot-deck step. The PRD model is fit using the family-level finer income category.

⁴⁶ There is one exception to the rule that provisional hot-deck steps involve univariate matching and univariate assignment. The provisional hot-deck step for cocaine and crack lifetime use utilizes multivariate assignment because both variables are used in the subsequent PRD model for heroin. The delta constraint refers to both predicted means, but in the calculation of Mahalanobis distance, only the cocaine predicted mean is used. Therefore, with respect to matching, this is not strictly univariate or multivariate; it is a little of both. See Section 6.5.1.1.4.

⁴⁷ There are exceptions. In a few imputation sets that use PMN Type 2 or PMN Type 3 (single RP/multiple PRD), provisional hot-deck steps are not completed because the variables earlier in the sequence are not used as covariates for variables later in the sequence. This occurs for some of the imputation sets for roster pairs (Section 14.5.4.3).

Figure 3.3 PMN Type 2: Multiple Response Propensity/Multiple Prediction



HD = hot deck; PRD = prediction; RP = response propensity.

In the multiple RP/multiple PRD type, multiple univariate PRD models are used. The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be computationally intensive due to the volume of model parameters and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, because of the obvious problem of specifying an accurate probability distribution underlying survey data, none of the existing solutions take the survey design into account. In the multiple RP/multiple PRD type, a multivariate model is fitted by a series of univariate parametric models (including the multinomial logistic case), such that variables modeled earlier in the sequence have a chance to be included in the covariate set for subsequent models in the sequence.

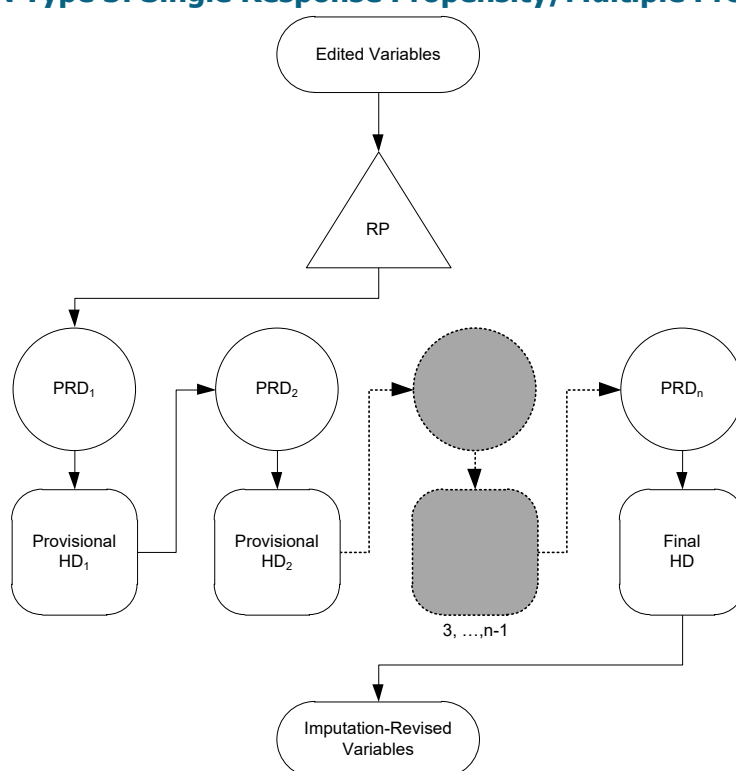
For variables imputed by PMN Type 2 and PMN Type 3 (single RP/multiple PRD), the order in which variables were modeled is important. Variables early in the sequence have the potential to be part of the set of covariates for variables later in the sequence, whereas variables later in the sequence cannot be used for modeling for the earlier variables because of missing values. Note that usually not all variables in the sequence were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For substances imputed using the PMN methodology, the sequence of imputation was determined by considering factors such as the level of prejudice and discrimination associated with the substances, the level of “missingness” in the data, and the

degree to which one set of substances could be used as predictors for other substances. The decisions on sequencing for other imputation sets were made using similar criteria. For some respondents, some but not all of the variables in the imputation set are missing. This gives rise to missingness patterns (Section 3.3.3.4). Typically, in the final hot-deck step, only the predictive mean vector elements corresponding to missing variables are used to match donors with item nonrespondents. However, likeness constraints (and sometimes logical constraints) are often used to preserve relationships between the missing and nonmissing variables. Although the nonmissing values would not be replaced by the corresponding values from the donor, some degree of correlation between missing and nonmissing variables is expected to be preserved using these constraints.

The multiple RP/multiple PRD type works well for closely related variables that have different domains and different nonresponse patterns, because the separate RP steps account for these. The recency and frequency variables provide a good example: the domain of the recency models for substances imputed using the PMN methodology consists of all lifetime users; the domain of the 12-month frequency model (if applicable) consists of all past year users; and the domain of the 30-day frequency model (if applicable) consists of all past month users. The provisional imputation-revised values may be used as covariates in later models, or even may be used to define the domains of later RP models.

3.4.3 Type 3: Single Response Propensity/Multiple Prediction

In PMN Type 3, the single RP/multiple PRD type, a single RP model is applied to all the variables modeled in the PRD steps. This is a less refined version of the preceding type, because it involves the fitting of only one RP model and is not as sensitive to different domains and response patterns among the outcome variables. The same weights are used for all PRD models. Among variables imputed using the PMN methodology, the lifetime substance use variables and source-of-income variables are examples of imputation sets that follow this type. The past month and past year marijuana ways of use variables are also imputed using this PMN type. However, provisional imputation is performed using colocated stochastic imputation (CSI) instead of hot-deck imputation. CSI is explained in detail in Section 3.6.3. [Figure 3.4](#) illustrates the single RP/multiple PRD type of PMN imputation.

Figure 3.4 PMN Type 3: Single Response Propensity/Multiple Prediction

HD = hot deck; PRD = prediction; RP = response propensity.

3.5 Development of the Modified Predictive Mean Neighborhood Methodology

Just as the PMN methodology was an attempt to improve the hot-deck imputation methods discussed in Section 3.1.1, the modPMN methodology is an attempt to improve PMN. The modPMN method is currently used for variables related to inhalants, hallucinogens, methamphetamine, prescription drugs (i.e., pain relievers, stimulants, sedatives, and tranquilizers), nicotine or tobacco vaping, flavoring vaping, synthetic marijuana use, synthetic stimulant use, kratom, illegally made fentanyl, GHB, nonprescription cough or cold medicines, and CBD or hemp. The modPMN method is also used for imputation of substance use disorder measures and variables related to adult mental health, adult depression, suicidal thoughts and behaviors among adults, receipt of treatment or services for use of alcohol or drugs, and receipt of treatment or services for mental health.

ModPMN retains the three basic steps of PMN—RP adjustment, PRD modeling, and imputation⁴⁸—but introduces some improvements into each of the three steps and includes a

⁴⁸ As opposed to PMN where the third step typically involves a hot-deck imputation, the imputation step under modPMN depends on the type of variable being imputed. Under modPMN, continuous variables are imputed using a hot-deck step, whereas categorical variables are imputed using a stochastic process. For additional details, see Section 3.6.3.

second cycle where each cycle looks much like PMN Type 2 (Section 3.4.2). The modPMN methodology addresses the following perceived shortcomings of PMN.

- Under modPMN, categorical variables are imputed stochastically using only the item nonrespondent's predicted values from the model. This CSI step for categorical variables replaces the hot-deck step of PMN, which does not "center" the imputed value. This decision was made in part due to the results described in Singh et al. (2004), where the notion of "Centered PMN" was introduced. Stochastic imputation is a much simpler way to center the imputed value of a categorical variable than hot-deck imputation, and it is implemented in a way that reduces the probability of extreme results (e.g., where either all item nonrespondents are imputed to "Yes" or all item nonrespondents are imputed to "No" for a dichotomous variable).
- For closely related variables, instead of using multivariate assignment in the imputation step (Section 3.3.3.3), closely related variables are imputed one at a time under modPMN, using a second cycle in a manner similar to IVEware (Raghunathan et al., 2002). The second cycle ensures that variables later in the sequence can be used to assist with imputation for variables earlier in the sequence. The one-at-a-time imputation approach greatly simplifies the predictive mean vectors and constraints that are used by PMN for hot-deck steps utilizing multivariate assignment.
- Instead of imputations being done separately within age groups (as under PMN), the domain members under modPMN are segregated into five bins based on their estimated propensity to respond, and the PRD modeling and imputation steps are implemented separately for each bin. This promotes "double protection," as described in Kott and Folsom (2010), without creating extreme weights, which can occur if item nonresponse is low. In contrast, the weight used in the hot-deck step of PMN does not promote double protection.
- Under modPMN, the segregation of substance use variables into imputation sets is done within measures and across substances instead of within substances and across measures (as under PMN). For example, instead of including hallucinogens recency, 12-month frequency, and 30-day frequency in a single imputation set, hallucinogens recency is in the same set as the recencies of inhalants, methamphetamine, and prescription psychotherapeutics. Similarly, substances are grouped within measures for imputation of 12-month frequency and 30-day frequency under modPMN. This simplifies the constraints and streamlines the logic.
- Under modPMN, child substance use measures (see Section 6.1.2) are handled after parent substance use measures have already undergone imputation. This also simplifies the constraints and logic and promotes a more customized approach to imputation within the small domains that are involved when child substance use measures are treated.
- Several warning messages produced by the modeling procedures in SUDAAN are disregarded when the PRD models are fit. This reduces the time required to fit the models with little impact on the predicted values.

The RP, PRD, and imputation steps are described in Section 3.6, and the arrangement of those steps into a procedure (which includes “cycling” as described above) is described in Section 3.7.

3.6 Procedural Steps for the Modified Predictive Mean Neighborhood Methodology

The implementation of modPMN on the NSDUH involves the same three basic steps that are used in PMN, as described in Section 3.3: RP adjustment, PRD modeling, and imputation. These steps are described in more detail in the following sections and are organized into blocks of 12 SAS programs to complete the modPMN imputation procedures as described in Section 3.7.⁴⁹ The descriptions below assume that the reader is already familiar with the material in Section 3.3, so the focus is mainly on the differences between these steps in modPMN and the analogous steps in PMN.

3.6.1 Step 1: Response Propensity Adjustment

The RP adjustment step under modPMN is very similar to the RP adjustment step under PMN as described in Section 3.3.1, with differences as follows:

- Under PMN, typically the RP step is run independently within age groups (see Section 3.8.1). Under modPMN, all age groups are included in the same RP model.
- Under PMN, each item respondent is assigned a weight adjustment factor: namely, the reciprocal of the estimated response propensity. This factor is multiplied by the analysis weight to create the weight used in the PRD model. Under modPMN, all domain members are divided into five “bins” of equal size based on their estimated response propensity, and the PRD modeling step and the imputation step are almost always completed within these bins.

These improvements are based on the “double protection” idea described in Kott and Folsom (2010). In PMN, the weights w_k used in the PRD models are

$$w_k = \frac{d_k}{p_k},$$

where w_k is the weight used in the PRD model for item respondent k , d_k is the analysis weight for item respondent k , and p_k is the estimated response propensity for item respondent k . When modPMN was first tested, as described in Chapter 5 of the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* report (Center for Behavioral Health

⁴⁹ For some variables, the imputation step of modPMN is completed in a single SAS program rather than a separate program for each bin, resulting in fewer than 12 programs being used. However, in these situations, the bin restrictions are still included in the imputation step. That is, stochastic imputation is still performed separately within each bin for categorical variables, and a likeness constraint is used in the hot deck step to restrict the donor pool based on the bin of the recipient for continuous variables.

Statistics and Quality, 2017), the weights described in Kott and Folsom (2010) were used in the PRD models:

$$w'_k = d_k \frac{1 - p_k}{p_k}.$$

However, this often resulted in extreme weights when the item response rate was high. This problem was addressed for modPMN by forming five RP bins instead of using the weights w'_k directly.

3.6.2 Step 2: Prediction Modeling

The PRD modeling step under modPMN is very similar to the PRD step under PMN as described in Section 3.3.2, with differences as follows:

- Under PMN, typically the PRD step is run independently within age groups (see Section 3.8.1). Under modPMN, the PRD step is run independently within the five RP bins. Section 3.6.1 explains how the five RP bins are formed.
- Under PMN, independent variables are dropped until the modeling procedure in SUDAAN produces no warning messages. Under modPMN, most of these warning messages are disregarded. As long as the algorithm converges, reasonable estimates of regression coefficients are produced, and reasonable test statistics associated with the regression coefficients are produced, the PRD model can be considered final.

[Table 3.6](#) is the modPMN analogue of [Table 3.2](#) for PMN. It shows which SUDAAN procedures are used to fit the PRD models for modPMN.

Table 3.6 Regression Models Used for Each Variable Imputed Using modPMN

Variable ¹	Domain	Type of Regression Model	SUDAAN [®] Procedure ²
Lifetime Drug Use: Inhalants, Hallucinogens, Methamphetamine, Nicotine Vaping, Flavoring Vaping, Synthetic Marijuana, Synthetic Stimulants, Kratom, Illegally Made Fentanyl, CBD or Hemp	All respondents	Binomial Logistic	RLOGIST
Recency of Any Drug Use: Prescription Drugs	All respondents	Multinomial Logistic	MULTILOG
Past Year Any Drug Use: OxyContin [®] , Fentanyl Products	All past year users of prescription pain relievers	Binomial Logistic	RLOGIST
Past Year Any Drug Use: Benzodiazepines	All past year users of prescription sedatives or tranquilizers	Binomial Logistic	RLOGIST

Table 3.6 Regression Models Used for Each Variable Imputed Using modPMN (continued)

Variable ¹	Domain	Type of Regression Model	SUDAAN® Procedure ²
Recency of Drug (Mis)use: Inhalants, Hallucinogens, Methamphetamine, Prescription Drugs, Nicotine Vaping, Flavoring Vaping, Synthetic Marijuana, Synthetic Stimulants, Kratom, Illegally Made Fentanyl, CBD or Hemp	All lifetime users/misusers	Multinomial Logistic	MULTILOG
Lifetime/Recency of Drug Use: GHB, Nonprescription Cough or Cold Medicines	All respondents	Multinomial Logistic	MULTILOG
Past Year Drug Misuse: OxyContin®, Benzodiazepines, Fentanyl Products	All past year users	Binomial Logistic	RLOGIST
12-Month Frequency of Drug Use: Inhalants, Hallucinogens, Methamphetamine	All past year users	Simple Linear	REGRESS
30-Day Frequency of Drug (Mis)use: Inhalants, Hallucinogens, Methamphetamine, Prescription Drugs, Nicotine Vaping	All past month users/misusers	Simple Linear	REGRESS
Age at First Drug Use: Inhalants, Hallucinogens, Methamphetamine, Nicotine Vaping	All lifetime users	Simple Linear	REGRESS
Past Year Initiation: Prescription Drugs	All past year misusers	Binomial Logistic	RLOGIST
Nicotine Dependence	All past month cigarette users	Binomial Logistic	RLOGIST
Past Year Substance Use Disorder: ³ Alcohol, Marijuana, Cocaine, Heroin, Inhalants, Hallucinogens, Methamphetamine, Prescription Drugs	All past year users	Binomial Logistic	RLOGIST
Past Year Substance Use Disorder Severity: ³ Alcohol, Marijuana, Cocaine, Heroin, Inhalants, Hallucinogens, Methamphetamine, Prescription Drugs	All past year users with a substance use disorder	Multinomial Logistic	MULTILOG

Table 3.6 Regression Models Used for Each Variable Imputed Using modPMN (continued)

Variable ¹	Domain	Type of Regression Model	SUDAAN® Procedure ²
Location or Type of Treatment or Services for Use of Alcohol or Drugs: Hospital as Inpatient, Residential Drug or Alcohol Treatment Center, Residential Mental Health Treatment Center, Other Inpatient Location, Outpatient Drug or Alcohol Treatment Center, Outpatient Mental Health Treatment Center, Office of Substance Use Treatment Professional, Doctor's Office, Hospital as Outpatient, School Counseling Center, Other Outpatient Location, Support Group, Emergency Room, Telehealth Treatment, Treatment in Jail or Prison, Peer Support Specialist	Lifetime drug or alcohol users	Binomial Logistic	RLOGIST
Location or Type of Treatment or Services for Use of Alcohol or Drugs: Detoxification Services	Past year drug or alcohol users	Binomial Logistic	RLOGIST
Location or Type of Treatment or Services for Use of Alcohol or Drugs: Prescription Medication for Alcohol Use	Lifetime alcohol users	Binomial Logistic	RLOGIST
Location or Type of Treatment or Services for Use of Alcohol or Drugs: Prescription Medication for Drug Use	Lifetime heroin or prescription pain reliever users	Binomial Logistic	RLOGIST
Adult Mental Health: Kessler-6, Past Month	All adults	Simple Linear	REGRESS
Adult Mental Health: Month in Past Year Worse Than Past 30 Days	All adults	Binomial Logistic	RLOGIST
Adult Mental Health: Kessler-6, Past Year	All adults with month in past year worse than past 30 days	Simple Linear	REGRESS
Adult Mental Health: WHODAS	Adults who reported experiencing at least one of the Kessler-6 symptoms at least a "little of the time"	Simple Linear (total WHODAS score); Cumulative Logit (alternative WHODAS score)	REGRESS; MULTILOG
Adult Depression: Lifetime MDE	All adults	Binomial Logistic	RLOGIST
Adult Depression: Past Year MDE	Adults with lifetime MDE	Binomial Logistic	RLOGIST
Adult Depression: Past Year MDE with Impairment	Adults with past year MDE	Binomial Logistic	RLOGIST
Adult Suicidal Thoughts, Plans, or Attempts in Past Year	All adults	Binomial Logistic	RLOGIST

Table 3.6 Regression Models Used for Each Variable Imputed Using modPMN (continued)

Variable ¹	Domain	Type of Regression Model	SUDAAN [®] Procedure ²
Adult Received Medical Attention for Suicide Attempt in Past Year	Adults who attempted suicide in past year	Binomial Logistic	RLOGIST
Adult Stayed Overnight in Hospital for Suicide Attempt in Past Year	Adults who received medical attention for suicide attempt in past year	Binomial Logistic	RLOGIST
Adult Suicidal Thoughts, Plans, or Attempts in Past Year due to COVID-19	Adults who had suicidal thoughts, plans, or attempts in past year	Binomial Logistic	RLOGIST
Location or Type of Treatment or Services for Mental Health: Hospital as Inpatient, Residential Mental Health Treatment Center, Residential Drug or Alcohol Treatment Center, Other Inpatient Location, Outpatient, Mental Health Treatment Center, Outpatient Drug or Alcohol Treatment Center, Office of Mental Health Professional, Doctor's Office, Hospital as Outpatient, School Counseling Center, Other Outpatient Location, Support Group, Emergency Room, Medication-Assisted Treatment, Telehealth Treatment, Treatment in Jail or Prison, Peer Support Specialist	All respondents	Binomial Logistic	RLOGIST

COVID-19 = coronavirus disease 2019; MDE = major depressive episode; modPMN = modified predictive mean neighborhood; WHODAS = World Health Organization Disability Assessment Schedule.

¹ The drug variables imputed using the modPMN methodology include those associated with the four prescription drug categories (prescription pain relievers, prescription tranquilizers, prescription stimulants, and prescription sedatives), hallucinogens, inhalants, methamphetamine, nicotine vaping, marijuana vaping, flavoring vaping, synthetic marijuana, synthetic stimulants, kratom, illegally made fentanyl, CBD or hemp, GHB, and nonprescription cough or cold medicines. Variables related to past year use and past year misuse of OxyContin[®], fentanyl products, and benzodiazepines are also imputed using modPMN.

² See RTI International (2013) for more information on all SAS[®]-callable SUDAAN[®] procedures in this table. SUDAAN is a registered trademark of Research Triangle Institute.

³ Past year substance use disorder and disorder severity are based on criteria from the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5).

3.6.3 Step 3: Imputation

For continuous variables, the imputation step of modPMN is identical to the imputation step of PMN (Section 3.3.3). The hot-deck step for modPMN almost always employs univariate matching and univariate assignment ([Table 3.4](#)).⁵⁰ However, for categorical variables, the

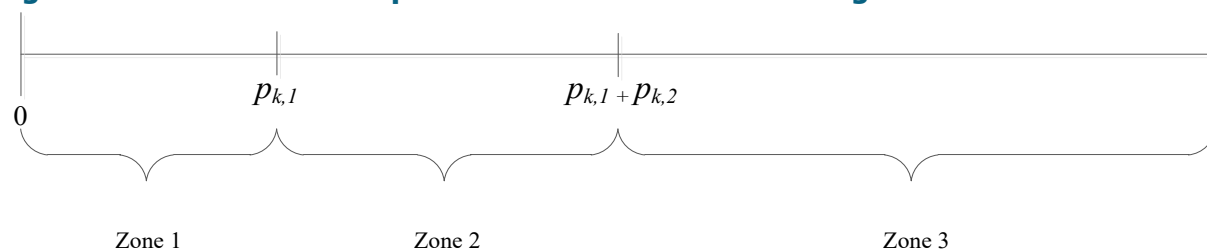
⁵⁰ The imputation sets that include the past month Kessler-6 variables, past year Kessler-6 variables, and WHODAS variables are an exception to this rule. For each of these sets, missing data for each of the variables in the set are replaced with values from a single donor.

imputation step of modPMN is not a hot-deck step. It is a random imputation based solely on the predicted means of the item nonrespondents. It also includes a feature that reduces the probability of extreme imputation results. The imputation step for categorical variables in modPMN is called "CSI." As for the PRD step, imputations are normally performed within the RP bin. For continuous variables, a likeness constraint is included that requires the donor to be in the same bin as the item nonrespondent. This section explains stochastic imputation first and then describes the collocation process in detail.

To implement stochastic (i.e., random) imputation in this context, a random number between 0 and 1 is drawn for each item nonrespondent, and the imputed value is equal to the predicted mean "zone," which contains the random number. In the simplest case, the variable to be imputed is dichotomous (0 = No and 1 = Yes), and if the predicted mean from the PRD step for a given item nonrespondent k is p_k and the random number drawn for item nonrespondent k is s_k , then the imputed value is 1 if, and only if, $s_k < p_k$. If the categorical variable to be imputed has G levels, then item nonrespondent k will have G predicted means from the PRD model such that $\sum_{g=1}^G p_{k,g} = 1$, where $p_{k,g}$ is the predicted probability that item nonrespondent k has a value of g .

Figure 3.5 shows stochastic imputation for a three-level variable that can take a value of 1, 2, or 3. If the random number s_k is in Zone 1 (i.e., $0 < s_k < p_{k,1}$), then the imputed value for item nonrespondent k is 1. If the random number s_k is in Zone 2 (i.e., $p_{k,1} < s_k < p_{k,1} + p_{k,2}$), then the imputed value for item nonrespondent k is 2. Finally, if the random number s_k is in Zone 3 (i.e., $p_{k,1} + p_{k,2} < s_k < 1$), then the imputed value for item nonrespondent k is 3.

Figure 3.5 Stochastic Imputation for a Three-Level Categorical Variable



Note that if the s_k values are drawn independently, then an extreme imputation can result by chance if all the s_k values happen to be close to 0 or close to 1, or if just by chance each s_k happens to fall into a narrow zone. The collocation process embedded in CSI is intended to reduce the probability of an extreme imputation by taking an approach more akin to systematic sampling than Poisson sampling, where each unit in the population is drawn into the sample resulting from independent decisions based on selection probability. In that case, the randomness of the process could lead to a very large sample or a very small sample.

A simple way to implement the systematic sampling-style approach for a dichotomous variable would be to randomly sort the R item nonrespondents (perhaps using s_k), choose f randomly from the uniform distribution on the interval $[0,1)$, and then impute k with a 1 if, and only if,

$$c_k = f / R + (k - 1) / R < p_k ,$$

where the value of k indicates the random sort order of the nonrespondents from 1 to R .

Observe that if all the p_k values were exactly equal to p , this approach would come very close to imputing a value of 1 ($p \times 100$) percent of the time. This suggests a slightly better methodology would replace c_k with

$$c'_k = \left(\sum_{j=1}^k w_j + [f - 1]w_k \right) / \sum_{j=1}^R w_j < p_k ,$$

where w_j is the weight of element j (and the elements were again randomly sorted and labeled in sort order). Here, if all the p_k values were exactly equal to p , this revised approach would produce a weighted average among the R elements needing imputation very close to p . Note that if all the w_j values were equal, c'_k would be the same as c_k .

If there are more than two categories in the variable needing imputation, the p_k value in the above method can be replaced by $p_{k,g}$, where $p_{k,g}$ is the probability that item nonrespondent k is in one of the first g categories (if there is no logical order for the categories, any order will do). Item nonrespondent k would be imputed into category g if, and only if,

$$p_{k,g-1} \leq c_k < p_{k,g} \text{ or } p_{k,g-1} \leq c'_k < p_{k,g} ,$$

where $p_{k,0}$ is defined to be 0. In fact, this latter approach involving the weights has been implemented on the NSDUH since 2015.

3.7 Implementation of the Modified Predictive Mean Neighborhood Methodology

Unlike the PMN methodology described in Section 3.4, there are not three types of modPMN. Under modPMN, most imputation sets in the 2023 NSDUH included more than one variable, and all were implemented using the same general arrangement of steps, somewhat akin to Type 2 of PMN (Section 3.4.2), but with a second cycle. Section 3.7.1 describes how variables are ordered in modPMN, and although the ordering is still important, the second cycle makes it less important for modPMN than it is for PMN. Section 3.7.2 describes the building blocks for modPMN implementation, a set of 12 SAS programs that implement the RP, PRD, and imputation steps for a single variable. This is similar to Type 1 of PMN (Section 3.4.1). Finally, Section 3.7.3 describes how the building blocks described in Section 3.7.2 are arranged to complete the imputation procedures for an imputation set.

Note that this section describes the implementation of modPMN in general terms. Readers interested in the implementation of modPMN for a specific imputation set are referred to the sections of the report covering that imputation set. The imputation sets that were processed with modPMN and the corresponding sections of the report are presented in [Table 3.7](#).

Table 3.7 Imputation Sets Processed Using modPMN

Set Number	Variables in Set	Section
1	Lifetime Substance Usage: Inhalants, Hallucinogens, Methamphetamine	6.5.2.3
2	Lifetime Substance Usage: Nicotine Vaping	6.5.2.3
3	Lifetime Substance Usage: Flavoring Vaping	8.6.6.3
4	Lifetime Substance Usage: Synthetic Marijuana, Synthetic Stimulants, Kratom, Illegally Made Fentanyl	8.6.6.3
5	Lifetime Substance Usage: CBD or Hemp	6.5.2.3
6	Recency of Any Use: Pain Relievers, Stimulants, Sedatives, Tranquilizers	6.5.2.3
7	Past Year Any Use: OxyContin®	6.5.2.3
8	Past Year Any Use: Fentanyl Products	6.5.2.3
9	Past Year Any Use: Benzodiazepines	6.5.2.3
10	Recency of (Mis)Use: Inhalants, Hallucinogens, Methamphetamine, Pain Relievers, Stimulants, Sedatives, Tranquilizers	6.5.2.3
11	Past Year Misuse: OxyContin®	6.5.2.3
12	Past Year Misuse: Fentanyl Products	6.5.2.3
13	Past Year Misuse: Benzodiazepines	6.5.2.3
14	Recency of Use: LSD, PCP, Ecstasy, Ketamine, DMT/AMT/Foxy, <i>Salvia divinorum</i>	6.5.2.3
15	Recency of Use: Nicotine Vaping	6.5.2.3
16	Recency of Use: Flavoring Vaping	8.6.6.3
17	Recency of Use: Synthetic Marijuana, Synthetic Stimulants, Kratom, Illegally Made Fentanyl	8.6.6.3
18	Lifetime Use and Recency of Use: GHB, Nonprescription Cough or Cold Medicines	7.2.6.3
19	Recency of Use: CBD or Hemp	6.5.2.3
20	12-month Frequency of Use: Inhalants, Hallucinogens, Methamphetamine	6.5.2.4
21	30-day Frequency of (Mis)Use: Inhalants, Hallucinogens, Methamphetamine, Pain Relievers, Stimulants, Sedatives, Tranquilizers	6.5.2.4
22	30-day Frequency of Use: Nicotine Vaping	6.5.2.4
23	Age at First Use: Inhalants, Hallucinogens, Methamphetamine	6.5.2.4
24	Age at First Use: LSD, PCP, Ecstasy	6.5.2.4
25	Age at First Use: Nicotine Vaping	6.5.2.4
26	Past Year Initiation of Drug Misuse: Pain Relievers, Stimulants, Sedatives, Tranquilizers	6.5.2.3
27	Nicotine Dependence	7.3.2.1
28	Past Year Substance Use Disorder: Alcohol	7.3.2.2
29	Past Year Substance Use Disorder: Marijuana	7.3.2.2
30	Past Year Substance Use Disorder: Hallucinogens	7.3.2.2
31	Past Year Substance Use Disorder: Inhalants	7.3.2.2
32	Past Year Substance Use Disorder: Cocaine, Stimulants, Methamphetamine	7.3.2.2

Table 3.7 Imputation Sets Processed Using modPMN (continued)

Set Number	Variables in Set	Section
33	Past Year Substance Use Disorder: Heroin and Pain Relievers	7.3.2.2
34	Past Year Substance Use Disorder: Sedatives, Tranquilizers	7.3.2.2
35	Past Year Substance Use Disorder: Past Year Users but Not Misusers of Prescription Drugs	7.3.2.2
36	Past Year Substance Use Disorder Severity: Alcohol	7.3.2.2
37	Past Year Substance Use Disorder Severity: Marijuana	7.3.2.2
38	Past Year Substance Use Disorder Severity: Hallucinogens	7.3.2.2
39	Past Year Substance Use Disorder Severity: Inhalants	7.3.2.2
40	Past Year Substance Use Disorder Severity: Cocaine, Stimulants, Methamphetamine	7.3.2.2
41	Past Year Substance Use Disorder Severity: Heroin and Pain Relievers	7.3.2.2
42	Past Year Substance Use Disorder Severity: Sedatives, Tranquilizers	7.3.2.2
43	Past Year Substance Use Disorder Severity: Past Year Users but Not Misusers of Prescription Drugs	7.3.2.2
44	Location or Type of Treatment or Services for Use of Alcohol or Drugs: Hospital as Inpatient, Residential Drug or Alcohol Treatment Center, Residential Mental Health Treatment Center, Other Inpatient Location, Outpatient Drug or Alcohol Treatment Center, Outpatient Mental Health Treatment Center, Office of Substance Use Treatment Professional, Doctor's Office, Hospital as Outpatient, School Counseling Center, Other Outpatient Location, Telehealth Treatment, Treatment in Jail or Prison, Prescription Medication for Alcohol Use, Prescription Medication for Drug Use	8.4.6
45	Location or Type of Treatment or Services for Use of Alcohol or Drugs: Support Group, Emergency Room, Peer Support Specialist, Detoxification Services	8.4.6
46	Adult Mental Health: Kessler-6, Past Month	10.7.2
47	Adult Mental Health: Month in Past Year Worse Than Past 30 Days	10.7.2
48	Adult Mental Health: Kessler-6, Past Year	10.7.2
49	Adult Mental Health: WHODAS	10.7.2
50	Adult Depression: Lifetime MDE	10.7.2
51	Adult Depression: Past Year MDE	10.7.2
52	Adult Depression: Past Year MDE with Impairment	10.7.2
53	Adult Suicidality: Thoughts, Plans, Attempts	10.7.2
54	Adult Suicidality: Medical Attention Received for Suicide Attempt in Past Year	10.7.2
55	Adult Suicidality: Overnight Stay in Hospital for Suicide Attempt in Past Year	10.7.2
56	Adult Suicidality: Thoughts, Plans, Attempts in Past Year due to COVID-19	10.7.2

Table 3.7 Imputation Sets Processed Using modPMN (continued)

Set Number	Variables in Set	Section
57	Location or Type of Treatment or Services for Mental Health: Hospital as Inpatient, Residential Mental Health Treatment Center, Residential Drug or Alcohol Treatment Center, Other Inpatient Location, Outpatient, Mental Health Treatment Center, Outpatient Drug or Alcohol Treatment Center, Office of Mental Health Professional, Doctor's Office, Hospital as Outpatient, School Counseling Center, Other Outpatient Location, Medication-Assisted Treatment, Telehealth Treatment, Treatment in Jail or Prison	10.7.1
58	Location or Type of Treatment or Services for Mental Health: Support Group, Emergency Room, Peer Support Specialist	10.7.1

COVID-19 = coronavirus disease 2019; MDE = major depressive episode; modPMN = modified predictive mean neighborhood; WHODAS = World Health Organization Disability Assessment Schedule.

3.7.1 Sequencing of Variables for modPMN

In PMN, the sequencing of variables within an imputation set is very important, because only variables earlier in the sequence are used as independent variables in the models for variables later in the sequence. Therefore, the imputation for variables later in the sequence can be said to be, in some sense, “better” than the imputation for variables earlier in the sequence. In the second cycle of modPMN, as described in Section 3.7.3, all other variables in the imputation set are used as independent variables for each variable. However, it may be said that the models are more fine-tuned for variables later in the sequence, because a larger set of independent variables was used for them in both the first cycle and the second cycle. The conclusion is that the sequencing of variables within a set is still important for modPMN but not quite as important as it is for PMN.

There are three clear candidates for the sequencing of variables within a set. The first is by descending item response rate, the method adopted by IVEware: all else being equal, imputation will have more of an impact on variables with lower item response rates. The second is based on the sequencing in the questionnaire, from beginning to end, which will tend to lead to the preservation of logical relationships and skip patterns. The third candidate approach is based on some measure of the variable’s analytic importance.

Not surprisingly, this third approach is highly subjective and difficult to determine. The first and second are straightforward to determine, but the first is more difficult to implement because it requires annual changes to the procedures, as item response rates tend to vary from year to year. (The sequencing of questions in the questionnaire rarely changes from one year to the next.) In consideration of the above, the approach taken is a compromise between these three approaches:

- For imputation sets 10 and 21 in [Table 3.7](#), the group containing inhalants, hallucinogens, and methamphetamine preceded all prescription drugs, but within these larger drug categories, sequencing was done by descending item response rate. [Table 3.8](#) provides an example of the sequencing.
- For imputation sets 53 and 56 in [Table 3.7](#), the order was static and imputed in the order shown in the table. This was done because most individuals who reported suicide

plans or a suicide attempt also reported having serious thoughts of suicide in the past year.

- For all other imputation sets in [Table 3.7](#), the variables were sequenced by descending item response rate.

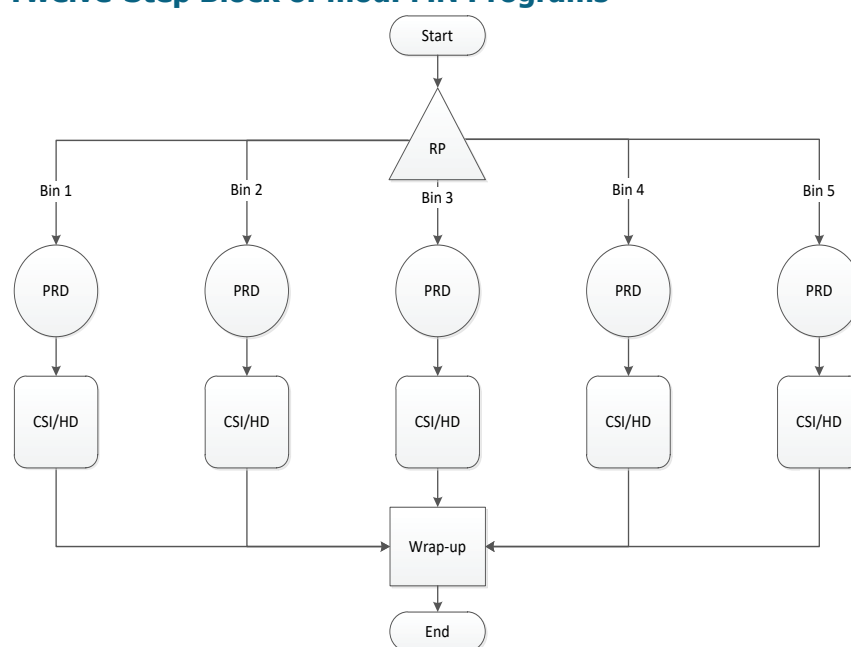
The rationale for the former is that it allows all inhalants, hallucinogens, and methamphetamine to be used as independent variables for the prescription drugs for both cycles. The prescription drugs were deemed more analytically important than the other drugs in the set.

Table 3.8 Example of Sequencing for Imputation Sets 10 and 21

Drug	Drug Type	Rank by Item Response Rate (Descending)	Position in Sequence
Hallucinogens	Nonprescription	1	1
Inhalants	Nonprescription	5	2
Methamphetamine	Nonprescription	7	3
Prescription Pain Relievers	Prescription	6	7
Prescription Tranquilizers	Prescription	3	5
Prescription Stimulants	Prescription	2	4
Prescription Sedatives	Prescription	4	6

3.7.2 Application of modPMN for a Single Variable

To apply modPMN for a single variable, 12 steps are used, as shown in [Figure 3.6](#). The first step implements the RP adjustment, which divides the domain members into five bins as described in Section 3.6.1. Within each bin, a PRD step (described in Section 3.6.2) and an imputation step (Section 3.6.3) are implemented. The imputation step involves CSI for categorical variables and hot-deck imputation for continuous variables. In the final step, the five output datasets from the within-bin imputation steps are merged into a single dataset, which contains imputed values for the variable that is undergoing imputation.

Figure 3.6 Twelve-Step Block of modPMN Programs

CSI = collocated stochastic imputation; HD = hot deck; modPMN = modified predictive mean neighborhood;
PRD = prediction; RP = response propensity.

3.7.3 Cycling in modPMN

When modPMN is applied to an imputation set, the three procedural steps of modPMN (Section 3.6) are always applied in the same coordinated way, by sequencing the 12-step blocks within two cycles across variables. At the end of each 12-step block in the first cycle, the variable has undergone provisional imputation, and at the end of each 12-step block in the second cycle, the variable has undergone final imputation. For each 12-step block, provisionally imputed values for all variables are used in the RP and PRD models whenever they are available.

[Table 3.9](#) lists the set of independent variables used in the RP and PRD models for each combination of cycle and variable number. Note that there is always a base set of independent variables, which are typically demographic variables that have either already undergone imputation or have no missing values for some other reason.

Table 3.9 Independent Variables Used in Response Propensity and Prediction Steps of modPMN

Cycle	Variable	Independent Variables
1	1	Base set
	2	Base set + provisionally imputed version of variable 1
	3	Base set + provisionally imputed versions of variables 1 and 2
	...	...
	k	Base set + provisionally imputed versions of variables 1, 2, ..., k - 1

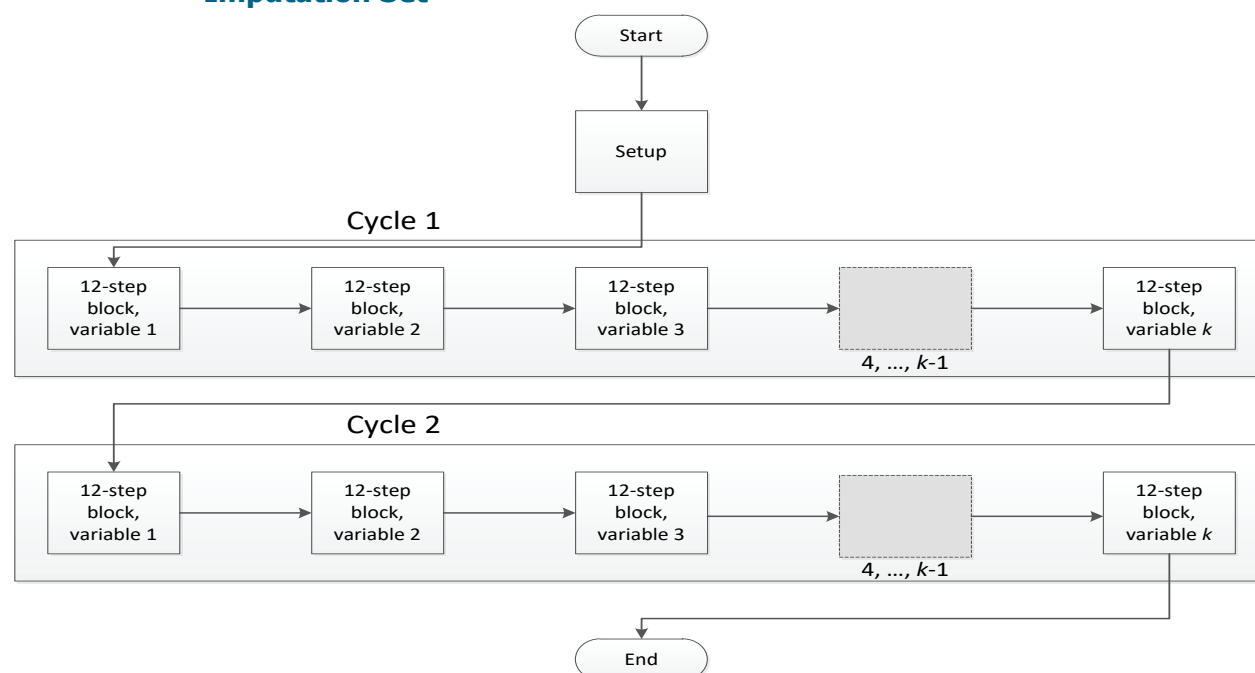
Table 3.9 Independent Variables Used in Response Propensity and Prediction Steps of modPMN (continued)

Cycle	Variable	Independent Variables
2	1	Base set + provisionally imputed versions of variables 2, 3, ..., k
	2	Base set + final imputed version of variable 1 + provisionally imputed versions of variables 3, 4, ..., k
	3	Base set + final imputed versions of variables 1 and 2 + provisionally imputed versions of variables 4, 5, ..., k
	...	...
	k	Base set + final imputed versions of variables 1, 2, ..., k - 1

modPMN = modified predictive mean neighborhood.

Figure 3.7 explains how the 12-step blocks are arranged in order to complete imputation for a particular imputation set with k variables. In the setup step, all variables are collected that are needed for the imputation of the set, including the edited versions of all variables undergoing imputation, the base set of covariates, the analysis weight, and the sample design variables needed for the PRD steps. The process then moves sequentially through the 12-step blocks for each variable nested within the two cycles. At the end of the second cycle, final imputed variables have been created for all variables in the imputation set.

Figure 3.7 Arranging the 12-Step Blocks to Complete Imputation for an Imputation Set



3.8 Special Auxiliary Variables: Age Group and State Rank

The age group and state of residence auxiliary variables apply to several of the imputation sets described in Chapters 4 through 8 and 10 through 14. Across variable groups, many imputation sets are processed separately by age group, regardless of the type of PMN that was used. The state of residence is used to construct a state-rank variable, which is then used in imputation for many of the variables that undergo imputation.

3.8.1 Age Groups

The variables related to substance use, household composition, income, and health insurance are highly correlated with age. This, along with the desire to use parallel processing to expedite the time it takes to impute all the variables, led to the decision to separate the imputation procedures for these variables into distinct age groups when using the PMN methodology. Therefore, the substance use variables for alcohol, cigarettes, cigars, pipes, smokeless tobacco, marijuana, marijuana ways of use, cocaine/crack, heroin, and the lifetime use variables for hallucinogen “child” drugs (i.e., LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, *Salvia divinorum*, and “any other” hallucinogen) were imputed within each of three age groups: 12 to 17, 18 to 25, and 26 or older. The household composition (roster-derived), income, and health insurance variables were imputed within the following four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older.⁵¹ The roster pair variables (i.e., the variables related to the relationship between two respondents from the same sampled dwelling unit) were divided into age groups depending on the ages of both pair members.

When imputation was performed using a hot-deck step, the age group restriction could be considered a likeness constraint. However, the models also were built separately within the age groups, so this restriction was not loosened unless no other options were available. Although the demographic variables did not always show a high correlation with age, the imputation of missing values in the demographic variables also was performed within age groups. This was done to maintain consistency with how the other variables were imputed and to facilitate parallel processing. The same three age groups that were used for substance use variables were also used for demographics. Occasionally, small sample sizes necessitated the aggregation of age groups at the modeling stage. In particular, the models for education level (highest grade completed) were fit within the 12-17 and 18-or-older age groups. In the employment status models, the 15-17 and 18-25 age groups were aggregated and models were fit within the 15-25 and 26-or-older age groups. Finally, all age groups were aggregated for the Hispanic/Latino group, marital status, and immigrant age-of-entry models.

⁵¹ Age groups were sometimes aggregated for the health insurance procedures. See Section 13.3 for details.

3.8.2 State Rank

Because state-level estimates are an important product of the NSDUH, there has been interest in requiring the donor to be from the same state as the recipient. However, this could not practically be implemented because of insufficient pools of donors.⁵² Therefore, information about the state of residence of each respondent was incorporated into the modeling and (when applicable) hot-deck steps of the PMN and modPMN procedures by grouping respondents into three categories based on the ranking of their state of residence with respect to some outcome. These state-rank variables were used as covariates in the RP and PRD steps and sometimes in likeness constraints in the hot-deck step.

3.9 Modeling Techniques to Improve Imputation Efficiency

The RP and PRD steps of PMN and modPMN can be time-consuming to implement because of the large number of models that must be fit each year, the long list of initial covariates used in these models, and the small domain sizes for some outcomes (e.g., frequency of heroin use in the past 30 days). To improve the efficiency of the PMN and modPMN imputation procedures, a more automated approach to fitting RP and PRD models was developed to reduce the amount of manual intervention required during the model-fitting process. For the 2023 NSDUH, the techniques described below were incorporated into the imputation procedures for the substance use variables; income variables; variables related to adult mental health, adult depression, and suicidal thoughts and behaviors among adults; mental health treatment; and substance use treatment. This more automated approach to fitting RP and PRD models will be integrated into the imputation procedures for other variables (e.g., health insurance) in future survey years as time and budget allow.

- For RP models that often require variables to be dropped, a smaller starting list of covariates was used, as described in the Evaluation of Imputation Methods report (Center for Behavioral Health Statistics and Quality, 2017). The starting and final covariate lists used in these models were examined over multiple survey years to identify covariates that were frequently dropped. Once identified, these covariates were then replaced with collapsed versions or removed from the starting list entirely.
- For both RP and PRD models, SAS code was developed to automatically detect and resolve data separation problems. Separation occurs in logistic regression when the cross-tabulation of a categorical response variable and a categorical covariate produces cells with no observations (Allison, 2008a). Covariates identified as causing a separation issue were collapsed, if possible, and rechecked. If it was not possible to collapse a covariate (e.g., gender), then it was automatically removed from the model. This process continued until the data separation issue was resolved or the covariate could no longer be collapsed, at which point it was removed from the model.

⁵² In the hot-deck step for some of the demographic variables, a likeness constraint required the donor to be from the same segment as the item nonrespondent. Segments never cross state lines, so this can be viewed as a refined use of the state of residence. In practice, this constraint often had to be removed because many segments included only a handful of unit respondents.

- For PRD models, covariates were automatically removed from the model when the overall Wald F test statistic for a related set of covariates was less than 1. Covariate sets were removed from the model one set at a time with interaction terms being dropped prior to removing main effects from the model.⁵³ This process was analogous to a backward selection stepwise regression, in that once a covariate was removed from the model, it remained excluded, and the process was repeated until all covariate sets in the model had a Wald F test statistic greater than or equal to 1.
- For variables imputed with modPMN, SAS code was developed to automatically collapse the five RP bins when the number of respondents within at least one bin was less than 10 and unlikely to support a large PRD model. When this occurred, a single PRD model was fit using respondents from all five RP bins combined.

⁵³ Additional restrictions were placed on the order in which variables were dropped for some models. For example, the age-at-first-use variable for a given drug would be dropped before removing indicators of lifetime use or most recent use of the same drug. Similarly, the 30-day frequency covariate for a given drug would be removed prior to removing the covariates associated with most recent use of the drug.

4. Editing and Imputation for the NSDUH Front-End Demographics Variables

4.1 Introduction

This chapter discusses procedures for editing and imputing the front-end demographic characteristics asked at the beginning of the survey. This includes age, birth date, sex assigned at birth, gender identity, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, U.S. military status, and education level (highest grade completed). Although the interview date was not classified as a demographic variable, it is used along with a respondent's birth date as the starting point for determining the respondent's current age. Therefore, the editing procedures for the interview date are included in this chapter. Likewise, though not specifically a demographic variable, a question on respondents' perceived health status is included in this section at the beginning of the interview and therefore is also discussed in this chapter. Finally, editing of interviewer debriefing variables corresponding to questions that interviewers completed after the in-person interview is discussed in this chapter. The set of demographic variables that are asked toward the end of the survey are discussed in Chapter 5.

For in-person interviews, questions concerning age, birth date, U.S. military status, and education level were asked by the interviewer, whereas for web interviews, these questions were self-administered. Questions concerning sex assigned at birth, gender identity, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and perceived health status were self-administered for all respondents beginning in 2023.

First, editing was performed on the variables. This editing could range from simply assigning legitimate skip codes (Section 2.4.2) to coding "OTHER, Specify" responses and resolving inconsistencies, as was the case for race and ethnicity (Section 2.3.1.2).

After editing, the variables were processed in one of four ways:

- No imputation required: interview date, age, and sex assigned at birth. These are described in Sections 4.2.1, 4.2.2, and 4.2.4.1, respectively. The edited data contained no missing values by design.
- No imputation performed: current state residency reported by interviewers (in-person interviews) or respondents (web-based interviews),⁵⁴ gender, number of times married, military status,⁵⁵ perceived health status, and interviewer debriefing (in-person

⁵⁴ The state residency that was reported by interviewers or respondents differs from the variable STATE, which contains the Federal Information Processing Standards (FIPS) codes for states and is created at the sampling stage. The variable STATE is recommended for use in state-level analyses.

⁵⁵ A final imputation-revised variable is not created for military service. However, a limited non-PMN imputation is performed for military service as part of the imputation procedures for health insurance, as described in Section 13.3.1.4.

interviews only).⁵⁶ These are described in Sections 4.2.3, 4.2.4.2, 4.2.6, 4.2.8, 4.2.10, and 4.2.11, respectively.

- Random assignment: birth date. This procedure is summarized in Section 4.2.5 because it is straightforward and does not involve the predictive mean neighborhood (PMN) method, which is described in Chapter 3.
- PMN: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level. These are described in Sections 4.2.6, 4.2.7, and 4.2.9.

Overall, the front-end demographics variables that are discussed in this chapter either did not have missing data by design or tended to have high item response rates. The item response rates when nonresponse was allowed (i.e., “don’t know” or “refused”) were more than 97 percent.

4.2 Editing Selected Front-End Demographics Variables

In this section, editing procedures are described for the following:

- interview date (Section 4.2.1);
- age (Section 4.2.2);
- state residency (Section 4.2.3);
- sex assigned at birth and gender identity (Section 4.2.4);
- birth date (Section 4.2.5);
- marital status and number of times married (if respondent had ever been married) (Section 4.2.6);
- race, Hispanic/Latino origin, and Hispanic/Latino group (Section 4.2.7);
- U.S. military service history and current military status (if respondents had ever been in the U.S. military) (Section 4.2.8);
- education level (Section 4.2.9);
- perceived health status (Section 4.2.10); and
- interviewer debriefing variables (Section 4.2.11).

4.2.1 Creating the Edited Interview Date Variable (INTDATE)

The program for the computer-assisted interviewing (CAI) version, as well as the web version, captured information about the time and date at the start of the interview, at the start of each section to which respondents were routed, and at the completion of the interview. This information was recorded in separate variables that were included in the interview data that were transmitted from the field to the data processing staff. These variables containing the start and end dates and times of the interview and sections are referred to as time and date “stamps.” For example, the time and date stamps for the beginning of the interview were captured in the variable TBEGINTR, the time and date stamps for the end of the interview were captured in the variable TENDINTR, and the time and date stamps for the beginning of the

⁵⁶ Interviewers complete the interviewer debriefing after the interview, but “interviewer debriefing” is included for completeness.

alcohol questionnaire section were captured in the variable TBEG_ALC.⁵⁷ These time and date stamps were sequentially arranged in the dataset in the order that the questionnaire sections appeared. In most cases, the time and date stamps were compared and edited to obtain the final edited interview date (INTDATE). The editing indicator (EIIDATE2) associated with the interview date variable specified the section date stamp that was used to create the edited interview date.⁵⁸

In some situations, the respondent did not finish the interview in one sitting and resumed the interview on another date, and the respondent's birthday occurred between the beginning and the end of the interview. In these situations, the interview date was set to the end-of-interview date stamp, which was consistent with the first date stamp after the respondent's birthday.

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

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

4.2.2 Creating the Edited Age Variable (AGE)

The age that respondents reported at the beginning of the interview (CALCAGE) was determined in one of two ways:

1. The respondents reported their birth date to the interviewer or in the web interview. This information was captured in the variable AGE1. For in-person respondents, the interviewer confirmed with the respondent that the birth date information had been recorded correctly and could not proceed further with the interview until the respondent verified the birth date. CALCAGE was determined from AGE1 and the date that was recorded at the start of the interview. For in-person respondents, the interviewer then confirmed with the respondent that this calculated age was correct. If the respondent indicated that the calculated age was not correct, then the interviewer could not proceed further until any additional corrections to the birth date had been made and the respondent verified both the date of birth as well as the calculated age. For web respondents, confirmation was required if the entered birth date made the calculated age greater than 120 years or less than 12 years.

⁵⁷ The CAI instrument includes time and date stamps for the start of individual questionnaire sections, but it does not include time and date stamps for the end of individual questionnaire sections because the last question in a section varies. For example, respondents who reported that they last used alcohol more than 12 months ago do not receive questions about alcohol use in the past 12 months or past 30 days.

⁵⁸ The 2023 questionnaire included a new self-administered front-end demographics section. This new section required some of the levels of the interview date indicator variables to be renumbered. Therefore, EIIATE was renamed EIIDATE2 in 2023.

2. If in-person respondents did not know or refused to report a birth date, the respondents were asked to provide their correct age, which was captured in the variable DKREFAGE. In this situation, CALCAGE was set to the age that had been recorded in DKREFAGE.

Interviews were terminated if age could not be determined or if the respondent age was less than 12 years old. Therefore, final respondents could not have missing data for their age.

Even after CALCAGE has been determined for respondents at the beginning of the interview, respondents could change their age in response to consistency checks in self-administered sections pertaining to their substance use (e.g., if they reported first using a drug at an age that was greater than their current age). Any changes that respondents made to their age during the questionnaire were captured in the temporary variable CURNTAGE. Therefore, it was possible for the age that was recorded for the respondent at the beginning of the interview (CALCAGE) to be different from the value of CURNTAGE that was captured at the end of the interview and stored in the variable NEWAGE.

The final age variable, AGE, was determined using CALCAGE and NEWAGE and three other sources: the age calculated from the final edited interview date (INTDATE) and the reported birth date (AGE1), the age corresponding to the “self” in the questionnaire household roster (if it existed) (Section 11.2.3), and the pre-interview age that was reported during screening of the dwelling unit (DU).⁵⁹ In most situations when the final edited continuous age was determined, priority was given to CALCAGE, NEWAGE, and the age that was calculated from AGE1 and INTDATE. If the final age (AGE) did not agree with the date of birth that had originally been entered (AGE1), the birth date also was edited. An intermediate value for age was determined by the logic listed in [Table 4.1](#). The rules are presented in the order in which they were applied.

⁵⁹ When contacting the DU for in-person interviews, the interviewer asked to speak with an adult resident of the household aged 18 or older who could serve as the screening respondent. Using a handheld computer, the interviewer completed a 5-minute procedure with the screening respondent that involved listing all household members along with their basic demographic data (including age). For web respondents, an adult resident of the DU who chose to participate could access the NSDUH web screening instrument from any device with Internet access (e.g., smartphone, tablet, computer). At the initial log-in page on the NSDUH survey website, an adult resident of the DU entered the participant code found on the lead letter or follow-up correspondence. Next, the screening respondent was asked to confirm that the respondent was an adult resident of the listed address for the DU and to read the informed consent text for the web-based screening. The respondent was then asked to provide basic demographic information for each household member aged 12 or older. The web-based screening program guided the respondents through the same roster questions included in the in-person screening program, with the question wording slightly altered to accommodate self-administration. Then for both in-person and web interviews, the computer used the demographic data in a preprogrammed selection algorithm to select zero, one, or two sample people, depending on the composition of the household.

Table 4.1 Rules to Create the Intermediate Value for Age

Intermediate Value for Age	Rule	Age Indicator
NEWAGE	NEWAGE = CALCAGE, and TBEG_TUT (the interview date stamp at the beginning of the tutorial) = INTDATE (the final edited interview date), or TBEG_TUT ≠ INTDATE but the respondent's birthday did not fall between TBEG_TUT and INTDATE, or TBEG_TUT ≠ INTDATE and the respondent's birthday does fall between TBEG_TUT and INTDATE but NEWAGE = age calculated from INTDATE and the reported birthdate	1
Age calculated from INTDATE and the reported birth date	TBEG_TUT ≠ INTDATE, and respondent's birthday falls between TBEG_TUT and INTDATE, and CALCAGE ≠ age calculated from INTDATE and the reported birth date, and age calculated from INTDATE and the reported birth date = screener age and/or roster age	2
NEWAGE	NEWAGE ≠ CALCAGE and NEWAGE = screener age and/or roster age, and date stamp at the beginning of the interview (TBEGINTR) was within the appropriate quarter	3
CALCAGE	NEWAGE ≠ CALCAGE and CALCAGE = screener age and/or roster age, and date stamp at the beginning of the interview (TBEGINTR) was within the appropriate quarter	4
Age calculated from INTDATE and the reported birth date	NEWAGE = CALCAGE and the year of TBEGINTR and TBEG_TUT are incorrect (EIIDATE = 5)	5
NEWAGE	NEWAGE ≠ CALCAGE and NEWAGE = roster age	6
CALCAGE	NEWAGE ≠ CALCAGE and CALCAGE = roster age	7
NEWAGE	NEWAGE ≠ age calculated from INTDATE and reported birth date but NEWAGE = CALCAGE, screener age, and roster age	8
CALCAGE	NEWAGE ≠ CALCAGE, but CALCAGE = age calculated from INTDATE and reported birth date CALCAGE is within 1 year of screener age and roster age	9

After these rules had been applied, this intermediate age value was compared with the age corresponding to the "self" in the household roster. For most respondents, the final edited value for the age variable (AGE) was set to this intermediate age value. However, there were exceptions for the in-person interviews as detailed below.

By the time the interviewer reached the roster part of the questionnaire, there had been multiple opportunities to change the value of age in response to consistency checks. This value of age was called CURNTAGE by the Blaise program. One of the consistency checks in the in-person household roster was to verify the value of the respondent's own entry for age in the

household roster (the “self” entry) against the value of CURNTAGE. If the self age differed from CURNTAGE, then the interviewer could either change the respondent’s age that was entered in the roster or override the consistency check and provide an explanation for why the roster age did not match CURNTAGE. If the consistency check for age was overridden, then the value for age corresponding to the “self” in the roster may not match the intermediate age value that was described previously. However, if sufficient explanation was given for overriding the age consistency check, other evidence pointed to the veracity of the roster age, and the difference between CURNTAGE and the roster age for self was at least 2 years, then AGE was set to the roster age, even if it disagreed with both NEWAGE and CALCAGE. For this to occur, all the following conditions had to be met:

1. The interviewer specifically indicated that the roster age was the correct one.
2. The pre-interview screener age matched the roster age.
3. If another household member also completed the interview, then the age from the household member’s roster must agree with the roster age.

It was possible for interviewers to jump back from the household roster or elsewhere in interviewer-administered sections (i.e., after respondents had completed the audio computer-assisted self-interviewing [ACASI] section of the interview) to change the respondent’s date of birth. In this situation, the values of CALCAGE and NEWAGE could be consistent, but these changes to the respondent’s age between the beginning and end of the interview would not be readily apparent. Rather, this pattern would become evident in reviewing ACASI data. For example, if respondents were aged 21 or older at the beginning of the interview and they did not change their age during the ACASI portion of the interview, questions were skipped in the consumption of alcohol section about alcohol use by people aged 12 to 20. However, if their final age indicated they were aged 12 to 20, then they would have been eligible to be asked questions about alcohol use by people aged 12 to 20⁶⁰ (see Sections 2.4.6 and 8.5).

Three age category variables were created from the final age for all respondents: CATAGE with four levels (12 to 17, 18 to 25, 26 to 34, and 35 or older); CATAG2 with three levels (12 to 17, 18 to 25, and 26 or older); and CATAG3 with five levels (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). These variables were used instead of the continuous age variables in some subsequent imputations and analyses.

4.2.3 Creating the Edited Current State Residency Variable (STATELOC)

Interviewers for in-person interviews or respondents for web interviews were instructed to use the “checkpoint” FIPE4 at the beginning of the interview to report the state where the DU was located. The term “checkpoint” refers to an item about the location of the DU or characteristics of the sample within the DU. These checkpoints were used to customize the wording of

⁶⁰ Web respondents may go backward in the instrument to change their age. However, to then go forward to where they left off would trigger any questions they had not previously received based on their original age. For example, if the respondent was aged 21 or older initially and then went back and changed to aged 20, the respondent would then be asked questions about alcohol use by people aged 12 to 20.

questions in subsequent sections of the interview, such as state-specific names for cash assistance from state or county welfare programs (Section 12.2.3).

The state entered in FIPE4 sometimes failed to match the state residence information that was used to sample a given respondent. Unless it could be verified as a valid mismatch, if the state information in FIPE4 was inconsistent with state information from the sample data, STATELOC was assigned a code for bad data (Section 2.4.3).⁶¹ In turn, the setting of STATELOC to bad data affected the editing of other variables that were dependent on the state where the respondent was reported to be a resident.

4.2.4 Creating the Edited Variables for Sex Assigned at Birth (IRSEX) and Gender Identity

Beginning in 2023, respondents were asked in question QD01 to report their sex assigned at birth on their original birth certificate. Previously, respondents were asked whether they were male or female. Despite this change in question wording, the variable names match those from prior years. Also beginning in 2023, question GI01 was added to ask respondents to report their current gender.

4.2.4.1 Sex Assigned at Birth

Respondents were required to enter a response in question QD01. Later they were asked to confirm their responses to sex assigned at birth and gender identity. Respondents who indicated that their original response was incorrect were asked in GICC02 to again report their sex assigned at birth. It was not possible to have missing values for sex assigned at birth. Although no imputation was required, the variable name IRSEX was used to maintain continuity with surveys prior to 2002 in which missing data could exist for QD01. It was not necessary to create an imputation indicator.

Similar to the situation described previously for age, interviewers also could jump back to the beginning of the interview and change the respondent's sex based on roster information. This type of change to QD01 between the beginning and end of the interview would not be readily apparent. However, this pattern would become evident in reviewing ACASI data. For example, if respondents were defined as being male at the beginning of the interview, questions in the self-administered healthcare section about pregnancy were skipped. However, if their final sex and age indicated that they were female and aged 12 to 44, then they would have been eligible to be asked questions about pregnancy (see Sections 2.4.6 and 10.2.1).

4.2.4.2 Gender Identity

Question GI01 asked respondents in an "enter all that apply" format to report their current gender. Respondents could select one or more of the response options (Male, Female,

⁶¹ Some of these mismatches existed for valid reasons, such as if a respondent had been selected in a DU in one state but had moved to another state. In these situations where field staff could investigate (i.e., in-person interviews), if FIPE4 reflected the state where the respondent was currently living, the edited variable STATELOC (corresponding to FIPE4) retained the value from FIPE4.

Transgender, I use a different term) and were also able to give a response of “don’t know” or “refuse.” Separate edited variables were created for each of the response options (GIDMALE, GIDFEMALE, GIDTRANS, and GIDOTHER, respectively). Respondents who selected “I use a different term” were asked in “OTHER, Specify” question GIOTH (edited variable GIOTHSP) to enter the term they use to describe their gender.

As mentioned in Section 4.2.4.1, respondents were asked follow-up questions to confirm their responses to QD01 and GI01. If respondents indicated that their original response to GI01 was correct, then responses to GI01 and GIOTH were used to create the edited variables. If respondents indicated that their original response to GI01 was not correct, then they were asked GICC03 (a repeat of GI01) and GICC04 (a repeat of GIOTH), where applicable, and these follow-up questions were used to create the edited variables.

Consistent with the procedures described in Section 2.4.5, if respondents did not report a particular gender identity in GI01 or GICC03, but they specified that gender identity in question GIOTH or GICC04, then the corresponding edited variable for that gender identity was assigned a code of 3 (Yes LOGICALLY ASSIGNED). For example, if respondents did not report being male in question GI01, but they specified in question GIOTH that they were male, then the corresponding edited variable GIDMALE was assigned a code of 3. However, if they specified a combination of listed gender identities (e.g., transgender, male), no edits were made to the specified gender identity variables.

4.2.5 Creating the Edited Birth Date Variable (BRTHDATE)

To proceed with the interview, respondents were required to provide their date of birth or current age (if they did not provide their date of birth) at the beginning of the interview. Each completed respondent possessed a current age, although some had missing birth dates. If the birth date was nonmissing but was inconsistent with AGE and INTDATE (either in the unedited data or as a result of editing age and/or interview date), then the reported birth month and day were preserved, and the birth year was logically edited according to the interview date and age.

If the birth date was missing, a birth date was randomly selected from all possible birth dates, given the final age and interview date. Each date in this period (365 or 366 days, depending on whether the period includes February 29 in a leap year) had an equal probability of selection. The editing indicator (EIBDATE) associated with the birth date specified how BRTHDATE was created.

4.2.6 Creating the Edited Marital Status Variables (MARITAL, EDMARIT, and NOMARR)

Question QD07 asked about respondents’ marital status if they were aged 15 or older. If respondents indicated that they were currently married, widowed, or divorced or separated, they also were asked to report the number of times they had been married (QD08).

MARITAL and NOMARR (corresponding to QD07 and QD08, respectively) underwent minimal processing. Legitimate skip codes were assigned to both variables (Section 2.4.2) if respondents were aged 12 to 14. Similarly, legitimate skip codes were assigned to NOMARR if respondents

had never been married. The base variable for creating an imputation-revised version of marital status was called EDMARIT. This variable was equivalent to MARITAL, with the exception that all legitimate skip codes were collapsed into a single legitimate skip code (99), and missing values were set to the SAS® missing code (.) so that they could be properly handled by the modeling programs.

4.2.7 Creating the Edited Race and Hispanic/Latino (Origin and Group) Variables

Two questions focused on the respondent's ethnicity⁶² (QD03 and QD04) and three focused on the respondent's race (QD05, QD05PI, and QD05ASIA). For those questions with multiple groups (QD04, QD05, QD05PI, and QD05ASIA), the respondent had the opportunity to select more than one group. For those questions with an "other" group (QD04, QD05PI, and QD05ASIA), respondents were asked to specify which other group best described them. QD05 did not have an "OTHER, Specify" response; however, if the respondents answered "don't know" to QD05, then they were prompted to specify which racial group described them.

As stated in the guidelines from the Office of Management and Budget,⁶³ "Hispanic/Latino" was categorized as an ethnicity, not a race. However, when given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group, resulting in missing data for the race question. Even though the final drug use tables were cross-classified with a variable that combined race and ethnicity, separate variables were initially created for race and ethnicity, and the race/ethnicity variables used in the tables were derived from these separate variables.

Due to the relationship between Hispanic/Latino origin and race reporting, Hispanic/Latino origin was used in the editing of race, and vice versa. In the process of editing race, the "OTHER, Specify" response to the Hispanic/Latino group question (QD04) was consulted (if it existed) if no race information was identified in QD05, QD05PI, or QD05ASIA. Similarly, in the process of editing the Hispanic/Latino group, the "OTHER, Specify" responses to the race questions (QD05, QD05PI, and QD05ASIA) were consulted (if they existed) if no Hispanic/Latino group information was identified in QD04. Because of the interdependence of race and Hispanic/Latino origin, the editing of these variables is discussed together in this section.

⁶² The questions about ethnicity were limited to determining whether a respondent was Hispanic/Latino or not, and the specific Hispanic/Latino group to which a Hispanic/Latino respondent belonged.

⁶³ The Office of Management and Budget released a notice, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" (Office of Management and Budget, 1997), that provides standards for maintaining, collecting, and presenting Federal data on race and ethnicity. These standards were applied to the 2023 NSDUH. On March 28, 2024, the Office of Management and Budget released updated standards in its review of Statistical Policy Directive No. 15 (SPD 15). Efforts are underway to comply with these new standards in future NSDUHs.

4.2.7.1 Groups Used in Race and Hispanic/Latino Variables

4.2.7.1.1 Race Groups

For editing purposes, 12 racial groups were created from the three race questions: 3 groups from QD05 (American Indian/Alaska Native, Black/African American, and White), 3 groups from QD05PI (Guamanian or Chamorro, Native Hawaiian, and Samoan), and 6 groups from QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese). Four additional groups were created from the "OTHER, Specify" responses: "Other Pacific Islander" (where the responses to QD05PI did not fit into any of the specific Pacific Islander groups), "Pacific Islander nonspecific" (where no response was selected to QD05PI, even though Hawaiian/Other Pacific Islander was selected in QD05), "Other Asian" (where the responses to QD05ASIA did not fit into any of the specific Asian groups), and "Asian nonspecific" (where no response was selected to QD05ASIA, even though Asian was selected in QD05). Respondents could be included in almost any subset of these groups. The only subsets that were not logically possible were those that included "Pacific Islander nonspecific" in combination with one or more specific Pacific Islander groups or "Asian nonspecific" in combination with one or more specific Asian groups. Combining the information from QD05, QD05PI, and QD05ASIA, as well as QD04 when necessary, allowed for the creation of all the edited and imputation-revised race variables.

4.2.7.1.2 Hispanic/Latino Groups

Respondents were given the choice of seven groups in QD04, and they could choose more than one group: Mexican/Mexican American/Mexicano/Chicano, Puerto Rican, Central or South American, Cuban/Cuban American, Dominican (from Dominican Republic), Spanish (from Spain), or some other Hispanic/Latino group.⁶⁴ If the respondents selected "Other," they were asked to specify which Hispanic group described them. This "Other, Specify" response would be either coded to a subset of the existing seven groups or set to missing. The "OTHER, Specify" responses to QD05, QD05PI, and/or QD05ASIA, if nonmissing, were consulted if no Hispanic/Latino origin group information was available from QD04. The final imputation-revised Hispanic/Latino group variable, IRHOGRP4, included all seven Hispanic/Latino group levels and a legitimate skip code (99) for respondents who were not Hispanic/Latino.

4.2.7.2 Classification of "OTHER, Specify" Codes

All "OTHER, Specify" responses from QD04, QD05, QD05PI, and QD05ASIA were assigned both a race code and a Hispanic/Latino code. A summary of how sets of "OTHER, Specify" codes were handled is provided in the following sections. Appendix C provides details on how each of the individual "OTHER, Specify" codes were handled.

4.2.7.2.1 Mapping of Race "OTHER, Specify" Codes

Each of the race codes fell into one of four categories: (1) codes that could be mapped directly to one or more of the race groups described in Section 4.2.7, (2) codes that could be indirectly

⁶⁴ When listing the six Hispanic/Latino defined categories in QD04, they shall henceforth be listed in this chapter as Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spanish.

mapped to one or more race groups, (3) codes that were informative in the final imputation, or (4) codes that were not informative for imputation. The edits resulting from codes that could be mapped to a race group were considered “final” because no imputation was necessary for them.

Directly mapped codes corresponded to one of the offered race groups from the questionnaire. Such responses include the race group itself or a geographic category code, which could be a racially homogenous country. The mapping of responses that were race groups did not usually depend on the question the response was from. A response of “Hawaiian” in any of questions QD05, QD05PI, or QD05Asia would be mapped to Hawaiian. Similarly, a response of an Asian or Pacific Islander country would not depend on the question. However, the mapping of responses that were non-Asian or non-Pacific Islander countries depended on which question the response was from. For example, a response of “Polish” to QD05 mapped to White because the Polish census data indicated nearly all Poles were White. Yet a response of “Polish” mapped to Other Asian or Other Pacific Islander if the response was from question QD05Asia or QD05PI, respectively.

Indirectly mapped codes resulted from an “OTHER, Specify” response, which stated a racially heterogeneous country. In these situations, the country’s census⁶⁵ was used to assign a race group. A random number was generated and compared with the proportions of races in the country’s census. For example, an entry of “Bolivian” would have a 70 percent chance of being allocated to the White and American Indian/Alaska Native groups, because the Bolivian census indicated 70 percent of Bolivians were Mestizo. If more than one country was given, races were assigned based on the proportions from each country and the results were combined. This method was also question dependent. Write-in responses to QD05PI or QD05Asia were mapped directly to the “Other Pacific Islander” or “Other Asian” racial group, respectively, unless a predominant Asian group existed within the country, in which case the response was directly mapped to that group. For instance, responses of “United Kingdom” to QD05Asia were mapped to “Asian Indian.”

“OTHER, Specify” responses that could not be mapped definitively to a specific race group resulted in incomplete values requiring imputation. These responses were considered either informative or noninformative for imputation. A response such as “mixed” was considered informative and resulted in an imputation among donors with two or more races. Logical constraints (see Section 3.3.3.1) were used to restrict the set of donors to ensure imputation-revised values were consistent with the information provided by the respondent. A response such as “American” was considered noninformative and resulted in an unrestricted imputation.

4.2.7.2.2 Mapping of Hispanic/Latino “OTHER, Specify” Codes

Certain Hispanic/Latino codes were considered “Definitely Hispanic.” If any of these appeared in QD05, QD05PI, or QD05ASIA, the respondent was considered Hispanic/Latino regardless of the response to QD03. Examples included “Hispanic” and “Dominicano” (Spanish for “Dominican”). There was also a code to handle respondents who were definitely not Hispanic/Latino (i.e., the

⁶⁵ See [The World Factbook](#) for more information.

respondent reported “Not Hispanic/Latino”). If this code appeared in QD04, QD05, QD05PI, or QD05ASIA, then the respondent was considered non-Hispanic/Latino regardless of the response to QD03. All other Hispanic/Latino codes either mapped directly to one or more of the seven Hispanic/Latino groups or provided no new information (e.g., Hispanic).

4.2.7.3 Edited Race Variables

4.2.7.3.1 Individual Race Groups (EDrace1-EDrace16)

Edited variables were created that correspond to the 16 racial groups described in Section 4.2.7.1.1. These variables were called EDrace xx , where xx represented a number between 1 and 16, corresponding to each of the 16 groups. [Table 4.2](#) shows the level for each group.

Table 4.2 Edited Variable Name for Each Race Group

Variable Name	Race Group
EDrace1	White
EDrace2	Black/African American
EDrace3	American Indian/Alaska Native
EDrace4	Native Hawaiian
EDrace5	Guamanian or Chamorro
EDrace6	Samoan
EDrace7	Other Pacific Islander
EDrace8	Asian Indian
EDrace9	Chinese
EDrace10	Filipino
EDrace11	Japanese
EDrace12	Korean
EDrace13	Vietnamese
EDrace14	Other Asian
EDrace15	Asian Nonspecific
EDrace16	Pacific Islander Nonspecific

EDrace xx could take on values of 1 to 4.

EDrace xx =

- 1, if the level xx was selected by the respondent in QD05, QD05PI, or QD05ASIA; else
- 2, if the level xx was indicated by a directly mapped code from the “OTHER, Specify” response(s) to QD05, QD05PI, or QD05ASIA; else
- 3, if no EDrace xx variables had values of 1 or 2, and the level xx was indicated by a directly mapped code from the “OTHER, Specify” response(s) to QD04 (Hispanic/Latino status); else

- 4, if (a) no EDrace_{xx} variables had values of 1, 2, or 3, and (b) the level *xx* was indicated by an indirectly mapped code from the “OTHER, Specify” response(s) to QD04, QD05, QD05PI, and/or QD05ASIA; else
- missing.

EDrace15 (Asian nonspecific) and EDrace16 (Pacific Islander nonspecific) were a little different from the others in that they did not map to a specific group from QD05, QD05PI, or QD05ASIA. They were used mainly to preserve a response of “Asian” or “Native Hawaiian or Other Pacific Islander” to QD05, even if the respondent selected nothing in QD05ASIA or QD05PI, respectively. The value of EDrace15 or EDrace16 was set to 1 in this situation. They could also have values of 2, 3, or 4, depending on the “OTHER, Specify” codes.⁶⁶

4.2.7.3.2 *Broad Groups of Race (EDRACE)*

EDRACE is a 42-level variable that indicates which of the five broad racial groups (White, Black/African American, American Indian/Alaska Native, Asian, or Native Hawaiian/Other Pacific Islander) were identified in QD04, QD05, QD05PI, and QD05ASIA. If no race group was identified, then EDRACE indicates how the imputation should be restricted based on the race of the donor. The first 5 levels of EDRACE correspond to the broad racial groups listed in QD05. EDRACE levels 6 to 15 encompass all combinations where two broad race groups were indicated, levels 16 to 25 encompass all combinations where three broad race groups were indicated, levels 26 to 30 encompass all combinations where four broad race groups were indicated, and level 31 is where all five broad race groups were indicated. If none of the broad racial groups were identified in QD04, QD05, QD05PI, and/or QD05ASIA, but other information was given that could inform imputation, then EDRACE could take on one of the following values:

- 32 (multiple race, no other information), if an “OTHER, Specify” response such as “biracial” or “mixed” appeared in QD04, QD05, or QD05ASIA; else
- 33 (indicated race other than White, no other information), if an “OTHER, Specify” response such as “nonwhite,” “brown,” “tan,” or similar responses in Spanish appeared in QD04, QD05, or QD05ASIA; else
- 34 (White, or both White and American Indian/Alaska Native), if the random assignment of a census data code resulted in imputation restricted to donors who were either White, or both White and American Indian/Alaska Native; else
- 35 (not American Indian/Alaska Native, in part or in full), if the random assignment of a census data code resulted in imputation restricted to donors who were not American Indian/Alaska Native, in part or in full; else
- 36 (non-Hispanic Mexican), if “Mexican” was mentioned in the QD05, QD05PI, and/or QD05ASIA “OTHER, Specify” responses, but QD03 = 2; else
- 37 (non-Hispanic Cuban), if “Cuban” was mentioned in the QD05, QD05PI, and/or QD05ASIA “OTHER, Specify” responses, but QD03 = 2; else

⁶⁶ For example, in EDrace15, a value of 2 indicated that the respondent wrote “Asian” in the QD05 other-specify. A value of 3 indicated that the response was obtained from the other-specify part of the Hispanic/Latino group question (QD04). A value of 4 indicated that the respondent gave a country of origin as a response to QD05, and the census for that country had “Asian” as one of its categories.

- 38 (non-Hispanic Central or South American), if "Central or South American" was mentioned in the QD05, QD05PI, and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 39 (non-Hispanic Dominican), if "Dominican" was mentioned in the QD05, QD05PI, and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 40 (non-Hispanic Spanish), if "Spanish" was mentioned in the QD05, QD05PI, and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 41 (non-Hispanic Puerto Rican), if "Puerto Rican" was mentioned in the QD05, QD05PI, and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 42 (non-Hispanic Cuban and Puerto Rican), if "Cuban" and "Puerto Rican" were mentioned in the QD05, QD05PI, and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2.

If no race groups were identified and no informative information given, then EDRACE was left missing.

4.2.7.3.3 *Finer Groups of Race (EDNWRACE)*

EDNWRACE was a 19-level edited variable used as a base variable for the imputation-revised finer racial group variable IRDETRACE. It was created using the following rules, under three possible scenarios:

Scenario 1: If only one of EDrace1 through EDrace16 was nonmissing,

EDNWRACE =

- 17 (more than one race), if respondents indicated they were multiracial in one of the "OTHER, Specify" responses; else
- 18 (Asian nonspecific only), if EDrace15 was the nonmissing variable; else
- 19 (Pacific Islander nonspecific only), if EDrace16 was the nonmissing variable; else
- xx (one known racial group only), where EDracexx was the nonmissing variable out of EDrace1 through EDrace7; else
- xx + 1 (one known racial group only), where EDracexx was the nonmissing variable out of EDrace8 through EDrace14.

Scenario 2: If more than one of EDrace1 through EDrace16 was nonmissing,

EDNWRACE =

- 8 (Native Hawaiian/Other Pacific Islander multiple group), if all race groups selected were in EDrace4 through EDrace7 or EDrace16; else
- 16 (Asian multiple group), if all race groups selected were in EDrace8 through EDrace15; else
- 17 (more than one race).

Scenario 3: If all of EDrace1 through EDrace16 were missing,

EDNWRACE =

- 17 (more than one race), if EDRAACE = 32; else
- missing.

4.2.7.4 Edited Hispanic/Latino Variables

4.2.7.4.1 *Hispanic/Latino Indicator (EDHOIND)*

An edited Hispanic/Latino indicator, EDHOIND, was created using responses to QD03 and, in rare situations, the "OTHER, Specify" responses to QD04, QD05, QD05PI, and/or QD05ASIA. This indicator variable was created as follows:

EDHOIND =

- 1 (Hispanic/Latino), if QD03 = 1 and no "OTHER, Specify" response stated that the respondent was definitely not Hispanic/Latino, or if the "OTHER, Specify" response to QD05 or QD05ASIA indicated that the respondent was definitely Hispanic/Latino; else
- 2 (not Hispanic/Latino), if QD03 = 2 and no "OTHER, Specify" response stated that the respondent was definitely Hispanic/Latino, or if the "OTHER, Specify" response to QD04, QD05, QD05PI, and/or QD05ASIA indicated that the respondent was definitely not Hispanic/Latino; else
- missing.

The race "OTHER, Specify" responses, which were considered "definitely Hispanic/Latino," and the single Hispanic/Latino "OTHER, Specify" response, which was considered "definitely not Hispanic/Latino," are listed in Appendix C.

4.2.7.4.2 *Individual Hispanic/Latino Group Groups (EDQD041-EDQD047)*

The edited variables EDQD041 through EDQD047 were created to match the seven Hispanic/Latino groups described in Section 4.2.7.1.2: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spanish, and Other Hispanic/Latino.

EDQD04xx =

- 1, if the level xx was selected by the respondent in QD04; else
- 2, if the "OTHER, Specify" response from QD04 mapped directly to level xx; else
- 3, if no EDQD04xx variables had values of 1 or 2, and the "OTHER, Specify" response from QD05 or QD05ASIA mapped directly to level xx; else
- missing.

4.2.7.4.3 *Edited Hispanic/Latino Group (EDHOGRP)*

The edited variable EDHOGRP was the base variable for creating the legacy imputation-revised Hispanic/Latino group variable IRHOGRP4, which does not have a "more than one Hispanic group" category. It had seven levels to match the seven Hispanic/Latino groups described in

Section 4.2.7.1.2, plus several other more general Hispanic/Latino levels that could be used as logical constraints on the set of potential donors (see Section 3.3.3.1). Those respondents with EDHOIND = 2 were assigned EDHOGRP = 99. It was created using the following rules, under four possible scenarios:

Scenario 1: If EDHOIND = 2,

EDHOGRP = 99.

Scenario 2: If EDHOIND = 1 or missing and only one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP = *xx*, where EDQD04*xx* was the nonmissing one.

Scenario 3: If EDHOIND = 1 or missing and more than one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP =

- 1 (Mexican), if EDQD041 was nonmissing; else
- 2 (Puerto Rican), if EDQD042 was nonmissing; else
- 3 (Central or South American), if EDQD043 was nonmissing; else
- 4 (Cuban), if EDQD044 was nonmissing; else
- 5 (Dominican), if EDQD045 was nonmissing; else
- 6 (Spanish), if EDQD046 was nonmissing; else
- 7 (Other), if EDQD047 was nonmissing.

For the multiple Hispanic/Latino group respondents, a priority rule was applied in determining a single Hispanic/Latino group in the following order: Mexican, Cuban, Puerto Rican, Central or South American, Dominican, Spanish, and Other Hispanic/Latino.

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

EDHOGRP =

- EDRACE + 7 (imputation restricted by race), if $1 \leq \text{EDRACE} \leq 30$; else
- missing.

4.2.8 Creating the Edited Military Service History and Military Veteran Status Variables

4.2.8.1 Current Service Variables (SERVICE and MILSTAT)

SERVICE and MILSTAT were created from questions QD09 and QD10, respectively, which asked respondents aged 17 or older about their current and past service in the U.S. military. Respondents who indicated in QD09 having never been in the U.S. armed forces were skipped out of QD10, which asked whether they were currently on active duty. Respondents who were currently on active duty in the U.S. military were not eligible to be interviewed for NSDUH.

Unlike the situation in most places in the interview (Section 2.4.2), responses of “don’t know” or “refused” to the question about lifetime military service (QD09) were treated as potential indications of military service. Thus, respondents who did not know or refused to report whether they had ever been in the U.S. military also were asked QD10 to determine their eligibility status for the interview.

Legitimate skip codes were assigned to SERVICE and MILSTAT if respondents were aged 12 to 16 (see Section 2.4.2). In addition, legitimate skip codes were assigned to MILSTAT if respondents were aged 17 or older but they reported in QD09 that they had never been in the U.S. armed forces.

If respondents reported in QD10 that they were currently on active military duty, they were asked a follow-up question to confirm this answer. The interview was terminated if respondents confirmed that they were on active duty in the U.S. military. Consequently, there were no final respondents in the final NSDUH data who reported that they currently were on active military duty. However, final respondents could be civilians who were in the military reserves at the time of the survey or were separated or retired from the military.

An additional question about full-time reserve status, V2B (edited variable MILRSRV), was administered if respondents reported in question QD10 that they were serving in a reserve component. Legitimate skip codes were assigned to MILRSRV if respondents were aged 12 to 16 or if respondents were aged 17 or older but they reported in QD09 that they had never been in the U.S. armed forces or they reported in QD10 that they were now separated or retired from the reserves or active duty.

4.2.8.2 Military Veteran Status Variables

Questions related to respondents’ military veteran status are included in the front-end demographics section for respondents who answered QD09 as “yes,” “don’t know,” or “refused” for whether they had ever served in the U.S. armed forces. These military veteran status questions pertained to whether respondents had ever served on active duty (QD10A), the period(s) when they served (QD10B1), and whether they served in a military combat zone where they drew imminent danger pay or hostile fire pay (QD10C).

Question QD10B1 was an “enter all that apply” question that allowed respondents to report more than one period for which they served. As noted in Section 2.4.4, separate edited variables were created for each response option. The response categories used for edited variables ACTD2001 through ACTD1941 have been coded as follows:

- 1 = Yes, and
- 2 = No (not entered).

The edited variables corresponding to questions QD10A through QD10C were assigned legitimate skip codes if respondents were aged 12 to 16 or indicated in question QD09 (SERVICE) that they had never served in the U.S. armed forces. Similarly, the edited variables for QD10B1 (ACTD2001 through ACTD1941) and QD10C (COMBATPY) were assigned legitimate skip codes if question QD10A (edited variable ACTDEVER) was answered as “no.”

In addition, the editing procedures included identifying inconsistencies between respondents' ages and reports of being on active duty in a particular period based on respondents' birth dates and the particular period when they reported being on active duty. Variables for the specific period of active duty were assigned codes for bad data (Section 2.4.3) if respondents' answers would lead to the conclusion that the respondents would have been younger than 16 or older than 65 during the entire period when they reported serving on active duty, including situations in which respondents reported service during a period before they were born. These edits were performed based on calculated upper and lower bounds for the possible ages when respondents could have been in the military during a given period of military service, based on their birthdates and the boundaries of the service periods in the questionnaire.

No editing was done if respondents would have been aged 16 to 65 at some point during the period for which they reported service.

4.2.9 Creating the Edited Highest Grade Completed Variables (EDUHIGHST and EDEDUC)

EDUHIGHST was created using the responses to the front-end education question QD11, which asked about the highest grade in school that the respondent had completed. This question is interview-administered for in-person respondents.

The CAI program compared respondents' highest level of education from QD11 and their current grade or year in school (if applicable) from question QD18, which is a self-administered question for both in-person and web respondents. A series of consistency checks were triggered based on certain answers in QD11 and QD18.

As an exception to the general principle of not editing across sections of the NSDUH interview (Section 2.4.1), editing to create EDUHIGHST considered data from consistency check question QD18CC04. An exception to this principle was made because respondents had the opportunity to revise their highest level of education in QD18CC04 or to replace their previous answer from QD11 with a missing value (i.e., "don't know" or "refused").

Specifically, if respondents changed their answer for their highest level of education in QD18CC04 to "don't know" or "refused," then the original answer from QD11 was replaced with the relevant code of 94 or 97, respectively, in EDUHIGHST (Section 2.2.1). Similarly, if QD18CC04 had a value from 0 to 20, then the default procedure was to assign the response from QD18CC04 to EDUHIGHST.

EDUHIGHST was assigned a code of 85 for bad data (Section 2.3) based on the conditions that are described below when respondents confirmed that their current grade in QD18 was the correct answer. Similar edits were implemented if respondents said that neither their current grade or year in school nor their highest level of education were correct, but the updated answer for the current grade and the highest level of education were still inconsistent.

- The respondents were currently in the 12th grade or lower and reported completing a grade or level of education in QD18CC04 that was higher than their current grade from QD18 or QD18CC03. These respondents logically would not have completed a higher grade than their current grade.
- The respondents were currently in the 12th grade or lower, and the highest completed grade from QD18CC04 was lower than the current grade by more than two grade levels. These responses would indicate that the respondents skipped from their previous grade level to the current grade level by more than two levels.
- The respondents were currently in the first year of college or higher and reported in QD18CC04 that they completed less than the 11th grade or that they completed the 12th grade but did not receive a high school diploma (QD18CC04 = 13). These responses indicate explicitly that respondents did not receive a high school diploma (when QD18CC04 = 13) or are implausible for the highest level of education if respondents were currently at a level of schooling beyond high school.
- The respondents were currently in their second year of college or higher and reported that their highest level of education was the receipt of a general educational development (GED) certificate of high school equivalency (QD18CC04 = 14). These respondents logically would be expected not to go from the receipt of GED to being in the second year of college or higher.

No editing was done to EDUHIGHST if respondents reported in QD18 that they were in the 12th grade at the time of the survey, confirmed that their current grade was correct, and reported that they previously completed the 12th grade but did not receive a diploma (QD18CC04 = 13). In this situation, respondents were assumed to be repeating the 12th grade.

In addition, if respondents confirmed that their current grade from QD18 was correct and reported in QD18CC04 that they received a high school diploma and that they were currently beyond their second year of college, then EDUHIGHST was assigned a code of 35, where 35 = Some college, no degree LOGICALLY ASSIGNED. In this situation, the receipt of a high school diploma would technically have been correct. Because QD18CC04 was self-administered for both in-person and web interviews, however, it was assumed that the respondents incorrectly reported in QD18CC04 that their highest level of education was a high school diploma, when they may have had some further education beyond high school in order to be beyond their second year in college.

A second variable, EDUHGHEDT, also was created as part of the back-end demographics processing (Section 5.2.6.3) that took into account additional patterns in the data for QD11 and QD18. However, EDUHGHEDT was not used in any published estimates involving educational attainment.

A temporary variable called EDEDUC was created to be used in the imputation process. It was equivalent to EDUHIGHST, except that missing values that described the type of item nonresponse (i.e., "don't know" or "refused") were set to the SAS missing code (.) so that they were properly handled by the imputation-related modeling programs.

4.2.10 Creating the Perceived Health Status Variable (HEALTH)

HEALTH was created from question QD12, which asked the respondents to rate their health on a 5-point scale from excellent to poor. No processing of this variable was done beyond that of assigning the edited variable name HEALTH instead of QD12.

4.2.11 Creating the Interviewer Debriefing Variables

For in-person interviews, the interviewer debriefing section was completed by the interviewer after the interview to obtain information about the potential quality of the interview. This section included information about factors that might have affected the quality of the data, such as the degree of privacy in the interview setting. These questions were not read aloud to the respondent and were completed on a tablet following the conclusion of the interview rather than on a laptop computer at the end of the CAI instrument.

Only minimal processing was done to the data in this section. Specifically, unedited variables were replaced with final descriptive variable names (e.g., PRIVACY for the variable pertaining to the interviewer's indication of how private the interview was). Where relevant, variables also were assigned legitimate skip codes based on the routing logic in this section.

4.3 Imputation for Selected Front-End Demographics Variables

In this section, the imputation procedures applied to marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level are described. These variables comprised five independent manifestations of the single response propensity (RP)/single prediction (PRD) type of PMN described in Section 3.4.3. Each imputation set is discussed in a separate section below. For details on the covariates used in the RP and PRD models, see Appendix D. For details on the hot-deck step, such as missingness patterns and constraints, see Appendix E.

4.3.1 Marital Status (Imputation Set 1)

The first demographic variable that underwent imputation was the marital status variable. The four substantive levels of the imputation-revised variable IRMARIT matched the four answer categories of QD07 (i.e., married, widowed, divorced or separated, or never married). Respondents aged 12 to 14 were automatically assigned an IRMARIT value of 99, a "legitimate skip" code. Because this was the first variable to undergo imputation in each cycle, there were no imputation-revised variables to use as auxiliary variables.

In marital status imputation procedures, only one RP model and one PRD model were fit. Single models were used across all age groups to ensure adequate sample size for response categories that would be rare within certain age groups (e.g., the "widowed" category for younger age groups). To account for the correlation between age and marital status, AGE was used in both the RP and PRD model steps and in a likeness constraint in the hot-deck step.

4.3.1.1 Response Propensity Step

The RP model for imputation set 1 utilized the preliminary analysis weight, PANALWT. The marital status question QD07 was only asked of respondents aged 15 or older. Therefore, the domain contained unit respondents with AGE greater than or equal to 15. The creation of the AGE variable is described in Section 4.2.2. Unit respondents in the domain with nonmissing EDMARIT values were considered item respondents. The EDMARIT variable is described in Section 4.2.6.

4.3.1.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the marital status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN®.⁶⁷ The outcome variable had four levels, which mapped to the four answer categories of QD07. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent selected each of the four answer categories of QD07.

4.3.1.3 Hot-Deck Step

The predicted means from the PRD step play a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps for marital status were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This allowed parallel processing, which decreased the time required for implementation. No logical constraints were used, and the only likeness constraint other than the delta constraint involved the continuous AGE variable. The few unit nonrespondents requiring imputation for this variable were usually handled in the first attempt to find a donor, due to the mild set of constraints and large domain. The only imputation-revised variable created in the hot-deck step was IRMARIT.

4.3.2 Race Variables (Imputation Set 2)

As discussed in Section 4.2.7, race and Hispanic/Latino origin were closely related. Therefore, race was used in the imputation of Hispanic/Latino origin and Hispanic/Latino group, and Hispanic/Latino origin was used in the imputation of race. Since race underwent imputation first, imputation-revised versions of the Hispanic/Latino indicator and the Hispanic/Latino group were not available. This precluded their usage in race models. However, they were used extensively in constraints in the hot-deck step. The RP, PRD, and hot-deck steps were run separately within three age groups: 12 to 17, 18 to 25, and 26 or older.

The race questions had a 98 percent response rate. However, due to high item nonresponse among Hispanic/Latino respondents, almost all of the race nonrespondents reported being of Hispanic/Latino origin ([Table 4.3](#)). The likeness constraints involving Hispanic/Latino group strongly influenced the final imputed values.

⁶⁷ SUDAAN® is a registered trademark of Research Triangle Institute.

4.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain for the RP models included all unit respondents. Item respondents were those with EDRACE values from 1 to 31 and EDNWRACE values from 1 to 17.

4.3.2.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the race variables were modeled using polytomous logistic regression as implemented by the MULTLOG procedure in SUDAAN. The outcome variable was the six-level variable EDRACEFORMODEL, which had the following levels:

- 1 White Only
- 2 Black/African American Only
- 3 American Indian/Alaska Native Only
- 4 Native Hawaiian/Other Pacific Islander Only
- 5 Asian
- 6 Multiple Race

EDRACEFORMODEL is a recode of the variable EDRACE, described in Section 4.2.7.3.2:

EDRACEFORMODEL =

- EDRACE, if $1 \leq \text{EDRACE} \leq 5$; else
- 6, if $6 \leq \text{EDRACE} \leq 32$; else
- missing.

The six predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDRACEFORMODEL.

4.3.2.3 Hot-Deck Step

Each item nonrespondent in the hot-deck step was assigned one of 14 missingness patterns. Twelve of the missingness patterns, all rare, were set up to handle instances where something was known about the race groups such as “known to be Asian.” The 13th missingness pattern, by far the most common, handled cases where nothing was known about the race groups. Logical constraints applied when something was known about the race groups. Otherwise, when nothing was known, only likeness constraints were used. Sometimes, what was “known” about the race groups came from a random assignment for indirectly mapped codes, as described in Section 4.2.7.2.1.

Besides the segment and delta likeness constraints, the likeness constraints based on Hispanic/Latino group were important determinants of the final imputed value, because the majority of the item nonrespondents for race were Hispanic/Latino. In 2023, 20.71 percent of the overall respondent pool were of Hispanic/Latino origin, but 76.34 percent of the item nonrespondents for race were of Hispanic/Latino origin. [Table 4.3](#) reports the distribution of Hispanic/Latino groups among race item nonrespondents in 2023. About 51 percent of the Hispanic/Latino nonrespondents were Mexican only.

Table 4.3 Hispanic/Latino Status of Item Nonrespondents for Race

Hispanic/Latino Status	Item Nonrespondents for Race	
	Number	Percentage
Not Hispanic/Latino or Missing Hispanic/Latino Indicator	265	23.66
Hispanic/Latino	855	76.34
Mexican Only	571	50.98
Puerto Rican Only	60	5.36
Central/South American Only	77	6.88
Dominican Only	45	4.02
Other/Unknown	102	9.11
Total	1,120	100.00

Depending on the missingness pattern, the item nonrespondent received values from the selected donor for some subset of EDNWRACE, EDrace1-EDrace7, a collapsed version of EDrace4-EDrace7, EDrace16, and a collapsed version of EDrace8-EDrace15. The collapsed version of EDrace4-EDrace7 and EDrace16 is the minimum of these variables and is an indicator of whether the respondent was Native Hawaiian/Other Pacific Islander. Similarly, the collapsed version of EDrace8-EDrace15 is the minimum of these variables and is an indicator of whether the respondent was Asian. Most receive values for all variables. Item nonrespondents in missingness pattern 2 (known to be Asian but missing an Asian finer category) received values for only EDNWRACE, item nonrespondents in missingness pattern 3 (known to be multiple race, but no other information) received values for everything except EDNWRACE, and item nonrespondents in missingness pattern 14 (known to be Native Hawaiian/Other Pacific Islander but missing a finer category) received values for only EDNWRACE. The imputation-revised versions of these variables are shown in [Table 4.4](#).

Table 4.4 Edited Race Variables and Their Imputation-Revised Counterparts

Edited Race Variable	Imputation-Revised Race Variable
EDrace1	IRRACEWH
EDrace2	IRRACEBK
EDrace3	IRRACENA
EDrace4	IRRACENH
EDrace5	IRRACEG
EDrace6	IRRACESM
EDrace7	IRRACPI2
EDrace4-EDrace7, EDrace16 (collapsed)	IRRACPIANY
EDrace8-EDrace15 (collapsed)	IRRACEAS
EDNWRACE	IRDETRACE

Due to the strict constraints, the delta constraint had to be dropped sometimes. However, the likeness constraints related to Hispanic/Latino group were never dropped.

4.3.3 Hispanic/Latino Origin Variable (Imputation Set 3)

For the Hispanic/Latino origin indicator, the RP, PRD, and hot-deck steps were run separately within three age groups: 12 to 17, 18 to 25, and 26 or older. Details on the procedures are given in the next four sections. The base variable for imputation, EDHOIND, is described in Section 4.2.7.4.1. The item response rate for this variable was much higher than for race.

4.3.3.1 Response Propensity Step

The response propensity models for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for EDHOIND.

4.3.3.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the Hispanic/Latino origin indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.⁶⁸ The outcome variable was EDHOIND. The single predicted mean was the predicted probability that the respondent was of Hispanic/Latino origin.

4.3.3.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino origin indicator included a single predicted mean from the PRD step, no logical constraints, and only the segment and delta likeness constraints. EDHOIND is the base variable for imputation, and the imputation-revised version is called IRHOIND.

4.3.3.4 Recodes for Additional Race/Ethnicity Variables

The imputation-revised race (IRDETRACE) and imputation-revised Hispanic/Latino indicator (IRHOIND) variables were used to create several additional combined race/ethnicity variables. One of these (RACE4) was used in the subsequent processing of imputation-revised variables and had four levels: non-Hispanic/Latino White, non-Hispanic/Latino Black/African American, Hispanic/Latino, and non-Hispanic/Latino other/multiple race. The NEWRACE2 and NEWRACE3 variables also were created from IRDETRACE and IRHOIND and were used extensively in the production of the *Results from the 2023 National Survey on Drug Use and Health: Detailed Tables* (Center for Behavioral Health Statistics and Quality, 2024c).

4.3.4 Hispanic/Latino Group Variable (Imputation Set 4)

The edited variable EDHOGRP, described in Section 4.2.7.4.3, categorized Hispanic/Latino respondents into Hispanic/Latino groups. These groups were directly mapped to the same

⁶⁸ In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

groups in the imputation-revised variable, IRHOGRP4, which had eight possible values: Puerto Rican, Mexican, Cuban, Central or South American, Dominican, Spanish, Other Hispanic/Latino, and not Hispanic/Latino. The closely related imputation-revised variable IRHOGRPM was also created to identify respondents who selected more than one Hispanic/Latino group. Recall that a priority rule was used to assign a single group to multiple-group respondents in the creation of EDHOGRP (and therefore IRHOGRP4).

Imputations were not conducted separately within age groups. The Hispanic/Latino group variables were created only for respondents of Hispanic/Latino origin as defined by IRHOIND. This results in a small domain. The models were likely to be better when age groups were combined because (1) none of the response groups were sparsely populated and (2) sufficiently large donor pools were ensured in the hot-deck step.

4.3.4.1 Response Propensity Step

The response propensity models for imputation set 4 utilized the preliminary analysis weight, PANALWT. The domain indicator included all respondents of Hispanic/Latino origin as defined by IRHOIND. Item respondents were those with a nonmissing value for EDHOGRP who selected only a single Hispanic/Latino group. The multiple-group respondents whose EDHOGRP was assigned by the priority rule (Scenario 3, described in Section 4.2.7.4.3) were not used to fit the PRD model in the next step.

4.3.4.2 Prediction Step

Because the model would have been much more difficult to fit if all seven levels were used, the EDHOGRP variable was collapsed into a four-level categorical variable (EDHOGRP2). Those with an EDHOGRP level of "Not Hispanic" were not in the domain. [Table 4.5](#) shows the mapping of EDHOGRP levels to EDHOGRP2 levels. Using the adjusted weights that were outputs of the RP step, EDHOGRP2 was then modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDHOGRP2.

Table 4.5 Mapping of EDHOGRP Levels to EDHOGRP2 Levels

EDHOGRP (Base Variable)	EDHOGRP2 (Modeled Variable)
Mexican	Mexican
Puerto Rican	Puerto Rican
Central or South American	Other Hispanic/Latino
Cuban	Cuban
Dominican	Other Hispanic/Latino
Spanish	Other Hispanic/Latino
Other Hispanic/Latino	Other Hispanic/Latino
Not Hispanic	Not Hispanic

4.3.4.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino group variables was straightforward. Besides the segment and delta likeness constraints, the most notable feature was a likeness constraint involving race. A six-level race variable was used as a covariate in the RP and PRD models with the following levels: White Only, Black/African American Only, American Indian/Alaska Native Only, Native Hawaiian/Other Pacific Islander Only, Asian Only, and Multiple Race. To further exploit the relationship between race and Hispanic/Latino group, a likeness constraint required the donor's selected race(s) to match a subset of the racial groups mentioned by the recipient. The constraint did not apply if the recipient was an item nonrespondent for race.

IRHOGRP4 was the imputation-revised version of EDHOGRP. The other imputation-revised variable IRHOG RPM was set equal to 8 (more than one Hispanic/Latino group) if either the respondent reported more than one group or the donor for a particular item nonrespondent reported more than one group. Otherwise, IRHOG RPM was set equal to IRHOGRP4.

4.3.4.4 Recodes for Additional Analyses

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

4.3.5 Education Level Variable (Imputation Set 5)

Similar to the imputation-revised Hispanic/Latino group variable, the imputation-revised education level variable was categorical with numerous levels, and the response variable for the PRD model was collapsed into fewer levels for ease of modeling. In most years, there have been fewer than 150 missing responses for this variable, so the application of the method has tended to be straightforward. Two age groups were used for RP and PRD modeling: 12 to 17 and 18 or older. However, the hot-deck step was implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older.

4.3.5.1 Response Propensity Step

The response propensity models for imputation set 5 utilized the preliminary analysis weight, PANALWT. The domain indicator for each of the two RP models included all unit respondents. Item respondents were those with a nonmissing value for EDEDUC.

4.3.5.2 Prediction Step

EDEDUC was collapsed into fewer levels for modeling. The response variables were different for the two PRD models: the response variable for the 12-17 age group had five levels, and the one for the 18-or-older age group had four. The mapping of EDEDUC to the response variable RESPEDUC is shown in [Table 4.6](#).

Using the adjusted weights that were outputs of the RP step, both PRD models were fit using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The

predicted means matched the levels of the response variable, so there were five predicted means for the 12-17 hot-deck step and four for the 18-25 and 26-or-older hot-deck steps.

Table 4.6 Mapping of EDEDUC Levels to RESPEDUC Levels

EDEDUC (Base Variable)	RESPEDUC (Modeled Variable)	
	12-17	18+
No Schooling Completed	Less Than Elementary School	Less Than High School
1 st Grade Completed		
2 nd Grade Completed		
3 rd Grade Completed		
4 th Grade Completed		
5 th Grade Completed	Elementary School	
6 th Grade Completed		
7 th Grade Completed		
8 th Grade Completed	Middle School	
9 th Grade Completed		
10 th Grade Completed	Some High School	
11 th Grade Completed		
12 th Grade, No Diploma		
Regular High School Diploma	High School or Higher	High School
GED Certificate of High School Completion		Some College/Associate’s Degree
Some College Credit, but No Degree		
Associate’s Degree		
Bachelor’s Degree		College Graduate
Master’s Degree		
Doctorate Degree		
Professional Degree Beyond a Bachelor’s Degree		

4.3.5.3 Hot-Deck Step

The hot-deck step for the education level variable was implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older. The only base variable was EDEDUC, and the imputation-revised version was called IREDUHIGHST. Both variables are based on the detailed 18-level variable, as compared with the simplified RESPEDUC variable used in the RP and PRD steps. No logical constraints were required. In addition to the segment and delta likeness constraints, the third likeness constraint required the donor to be the same age as the recipient. This was an especially important constraint for the 12-17 age group, because the age covariate often had to be dropped from the PRD model due to near-empty cells when the variables were cross-tabulated, causing instability in the estimates.

4.3.5.4 Recode for Additional Education Variable

EDUHIGHCAT, a recoded education variable, was created using the imputation-revised highest grade completed variable (IREDUHIGHST). EDUHIGHCAT had five levels (less than high school and aged 18 or older, high school graduate and aged 18 or older, some college/Associate's Degree and aged 18 or older, college graduate and aged 18 or older, or 12 to 17 years old).

If a respondent was 12 to 17 years old, then EDUHIGHCAT was assigned that level. If IREDUHIGHST was "No Schooling Completed" or any grade level up to "12th Grade, No Diploma," then EDUHIGHCAT was designated "Less than High School." If IREDUHIGHST was either "High School Diploma" or "GED," then EDUHIGHCAT was assigned "High School." If IREDUHIGHST was "Some College Credit, but No Degree" or "Associate's Degree," then EDUHIGHCAT was designated "Some College/Associate's Degree." Any degree beyond "Associate's Degree" was categorized as "College Graduate." EDUHIGHCAT was often used as a covariate in later imputation models.

5. Editing and Imputation for the NSDUH Back-End Demographics Variables

5.1 Introduction

This chapter discusses procedures for editing and imputing the demographic variables in the 2023 National Survey on Drug Use and Health (NSDUH) created from questions asked at the end of the interview. These questions were self-administered for both web and in-person respondents. Editing procedures are described for the following demographics sections:

- moves in the past 12 months,
- immigrant status,
- sexual attraction and identity,
- English language proficiency,
- disability,
- current education,
- military families, and
- employment and workplace.

This chapter also discusses imputation procedures for the variables pertaining to immigrant status and current employment status. Other edited variables for self-administered demographics that are discussed in this chapter were not imputed.

5.2 Editing the Back-End Demographics Variables

This section documents the editing procedures for the demographics sections. As noted in Section 1.1, these demographics sections of the interview were administered by the interviewers using audio computer-assisted self-interviewing (ACASI) for in-person interviews. All sections of web-based interviews were self-administered.

5.2.1 Editing for Patterned Responses

As discussed in Section 2.3.2.2, edits for patterned responses were implemented in the back-end demographics questions. These edits resulted in data in back-end demographics variables being set to “bad data” (Section 2.3) for about 100 respondents in 2023 who keyed responses of only 1 (or variants such as 11) or 2 (or variants such as 22) wherever possible in the back-end demographics, education, and employment questions. This number of respondents with patterned responses in their back-end demographic data represented about 0.15 percent of final respondents.

5.2.2 Moves in the Past 12 Months

Respondents were asked in question QD13 (edited variable MOVSPNPYR) how many times in the past 12 months they had moved, including moves within their city/town and between cities/towns. If respondents moved at least once in the past 12 months, they were asked in question QD13a (edited variable MOVFRSTAT) to report the state where they were living a year

prior to their interview date. If respondents reported zero moves in the past 12 months in QD13, then MOVFRSTAT was assigned a legitimate skip code (Section 2.4.2).

5.2.3 Immigrant Status

Respondents were asked a series of questions pertaining to immigrant status, including whether they were born in the United States and, if not, the country or U.S. territory where they were born, as well as the length of time they have lived in the United States.

5.2.3.1 Births within or Outside of the United States

Respondents were asked in question QD15 to report the country or U.S. territory they were born in if they reported in QD14 that they were not born in the United States. Note that respondents who were born in a U.S. territory may have reported that they were born in the United States and would not be asked question QD15. Respondents who specified a state or the District of Columbia in QD15 were logically inferred to have been born in the United States. Respondents who specified in QD15 that they were born in a U.S. military installation outside of the United States were not logically inferred to have been born in the United States.

When respondents reported in QD14 that they were not born in the United States, but they specified a state or the District of Columbia in QD15, the edited variable IMMBORNUS corresponding to QD14 was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED. The data in QD15 (edited variable IMMBORNUS) were not overwritten with codes of 989 (LEGITIMATE SKIP Logically assigned; Section 2.4.2) to allow analysts to see the responses in QD15 that led to the logical inference that respondents were born in the United States.

Otherwise, when respondents reported in QD14 that they were born in the United States, IMMBORNUS was assigned a legitimate skip code. This assignment of a legitimate skip code did not include situations in which respondents had missing data for where they were born (i.e., don't know, refused, or bad data). However, if respondents refused to answer QD14, then that refusal was propagated to IMMBORNUS.

5.2.3.2 Length of Time Living in the United States

The logic for asking questions QD16a through QD16c about the length of time that respondents have lived in the United States (if QD14 was answered as "no") takes into account the number of moves in the past 12 months from QD13. Whether a respondent had lived in the United States for at least 1 year (QD16a) was asked only if respondents who were not born in the United States had reported at least one move in the past 12 months in question QD13 or if they answered QD13 as "don't know" or "refused." The assumption was that immigrants who reported no moves in the past 12 months logically have lived in the United States for at least 1 year. The number of years that immigrant respondents have lived in the United States (QD16b) was asked if they reported in QD16a that they had lived in the United States for at least 1 year or they reported no moves in the past 12 months in QD13.

Editing of the variables for the length of time that immigrant respondents had lived in the United States principally involved assigning legitimate skip codes (Section 2.4.2) when the

associated questions did not apply. In particular, if respondents reported in QD14 that they had been born in the United States (IMMBORNUS = 1), then the edited variables IMMYRMORE and IMMOSNUM (corresponding to QD16a and QD16c, respectively) were assigned a code of 99 and IMMYRSNUM (corresponding to QD16b) was assigned a code of 999. If the edits described in Section 5.2.3.1 resulted in respondents being logically inferred to have been born in the United States (IMMBORNUS = 3), then any nonblank values in IMMYRMORE and IMMOSNUM were overwritten with a code of 89 (LEGITIMATE SKIP Logically assigned) and nonblank values in IMMYRSNUM were overwritten with a code of 989.

Similarly, a legitimate skip code was assigned to IMMYRMORE and IMMOSNUM if immigrant respondents reported no moves in the past 12 months in QD13. If respondents were asked QD16a, then a legitimate skip code was assigned to IMMOSNUM if respondents reported in QD16a that they had lived in the United States for at least 1 year or to IMMYRSNUM if respondents reported that they had not lived in the United States for at least 1 year.

The number of years that respondents reported living in the United States in QD16b also was compared with respondents' ages. If the number of years in QD16b was greater than respondents' current ages, then IMMYRSNUM was assigned a code for bad data (Section 2.3) to indicate that the length of time that respondents reported living in the United States was inconsistent with their age. No editing was done to IMMYRSNUM if QD16b was equal to respondents' current ages.

5.2.4 Sexual Attraction and Identity

Questions QD62 (edited variable SEXATTRACT2) and QD63 (edited variable SEXIDENT2) asked all respondents about their sexual attraction and sexual identity, respectively. Several changes were made to these questions starting in 2023, and the edited variables used previously (SEXATTRACT and SEXIDENT) were renamed. First, both questions were newly asked of adolescents younger than 18, along with the adults aged 18 or older. Also, three additional response options for QD63 allowed respondents to report that they use a different term other than one listed, they are not sure about their sexual identity, or they do not know what the question is asking. Respondents who reported in QD63 using a different term were asked in new "OTHER, Specify" question QD63OTH (edited variable SEXIDOTHSP) to describe their sexual identity. When respondents reported in QD63OTH one of the response options offered in QD63 but did not select that response option in QD63, the edited variable SEXIDENT2 was given a logically assigned code. The logically assigned code was determined by the value from QD63 plus an added value of 10. For example, respondents who did not indicate QD63=1 (Heterosexual, that is, straight) but then specified in QD63OTH that they were heterosexual were assigned a code of SEXIDENT2=11 (Heterosexual, that is straight, LOGICALLY ASSIGNED).

QD62 People are different in their sexual attraction to other people. Which statement best describes your feelings?

[IF RSEX=5 (Male assigned at birth)]

- 1 I am only attracted to females
- 2 I am mostly attracted to females

- 3 I am equally attracted to females and males
- 4 I am mostly attracted to males
- 5 I am only attracted to males
- 6 I am not sure
- DK/REF

[IF RSEX=9 (Female assigned at birth)]

- 1 I am only attracted to males
- 2 I am mostly attracted to males
- 3 I am equally attracted to males and females
- 4 I am mostly attracted to females
- 5 I am only attracted to females
- 6 I am not sure

QD63 Which one of the following do you consider yourself to be?

- 1 Heterosexual, that is, straight
- 2 Gay or lesbian
- 3 Bisexual
- 4 I use a different term
- 5 I am not sure about my sexual identity
- 6 I do not know what this question is asking
- DK/REF

QD63OTH [IF QD63=4] What term do you use to describe your sexual identity

_____ [ALLOW 50 CHARACTERS]
DK/REF

As shown above, the order of the response options for question QD62 varied according to the sex assigned at birth of the respondent. However, interviewers could go back and change the response to the sex assigned at birth question (QD01) after in-person respondents had completed the ACASI portion of the interview. When this situation occurred, the computer-assisted interviewing program did not allow respondents to reanswer QD62 with the new response options corresponding to the revised gender (Section 2.4.6). For example, if a respondent's final sex assigned at birth was reported to be female, but the respondent was reported to be male at the time that the respondent was asked QD62, then the respondent was shown the levels that were administered to males. If the respondent actually was female and was attracted only to males, then the respondent would be expected to enter 5 to this "male" version of the question, for "I am only attracted to males." If the interviewer jumped back to the beginning of the interview and changed the respondent's final sex assigned at birth to be female, however, then the meaning of level 5 in the edited variable SEXATTRACT2 would become reinterpreted to mean that the respondent is attracted only to females, as per the descriptions shown above for QD62. No editing was done to SEXATTRACT2, other than to assign final descriptive names to the variable.

5.2.5 English Language Proficiency and Disability

Question QD55 measures English language proficiency by asking how well the respondent speaks English by the respondent's own account with four response options: very well, well, not well, and not at all. This item is intended to help monitor health disparities in populations with limited proficiency with English.

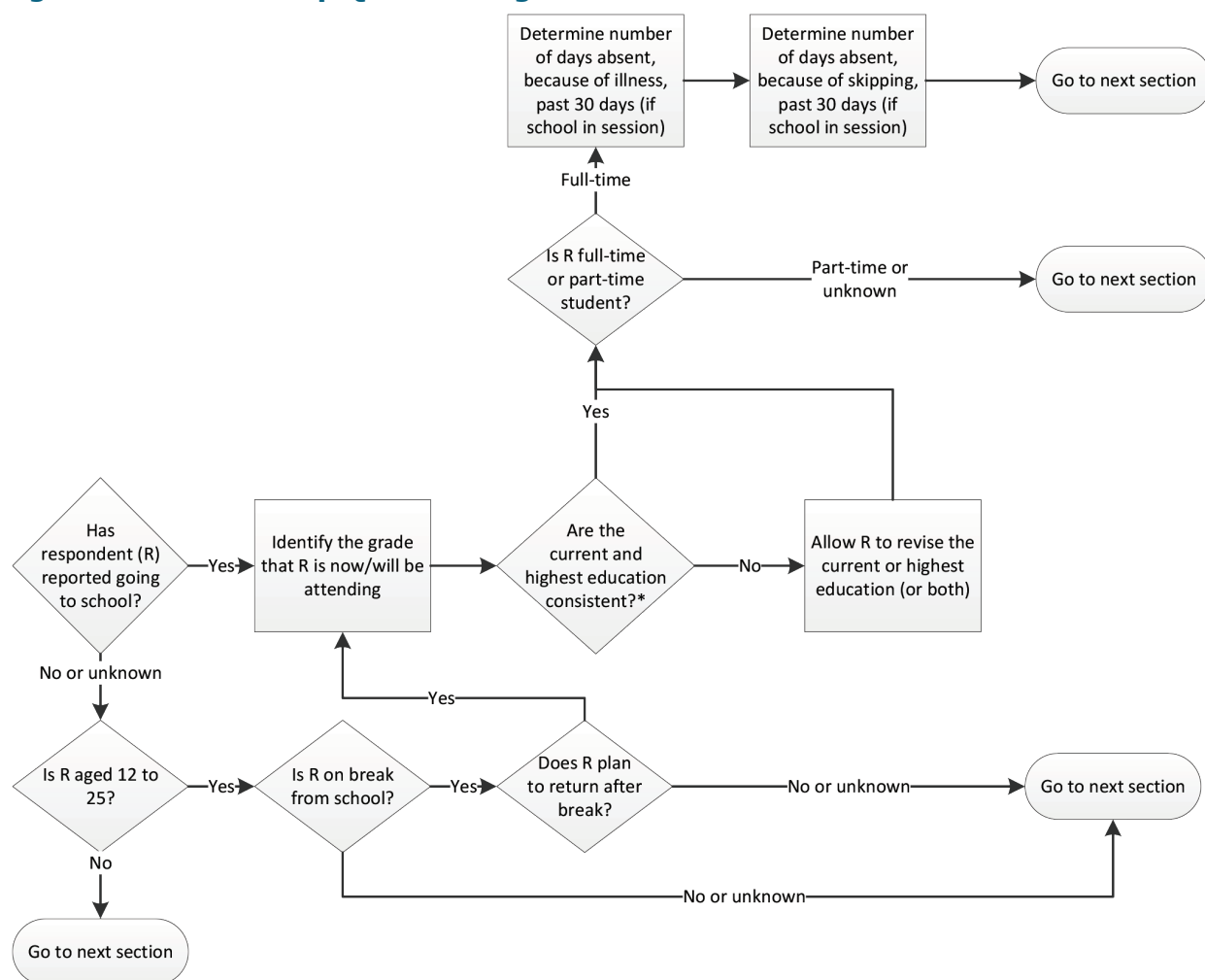
A series of six questions were used to measure functional disability (SEE, HEAR, WALK, MEM, CARE, and COMM), based on the [Washington Group Short Set on Functioning](#) . Each question has four response options asking about level of difficulty associated with each task: no difficulty, some difficulty, a lot of difficulty, and cannot do at all.

All respondents were asked these questions. No editing was done to the associated variables, other than to assign final descriptive names to the variables (e.g., SPEAKENGL, corresponding to question QD55 for how well respondents speak English).

5.2.6 Current Education

The questions for current education assessed whether respondents were currently enrolled in school. If so, respondents were asked to report their current grade in school (or the grade they would be in once they returned from school break), whether they were a full-time or part-time student, and if they were full-time students, the number of days that they missed school in the past 30 days because they were sick or because they skipped school (questions QD18 through QD21). [Figure 5.1](#) shows the logic for asking follow-up questions if respondents reported that they were enrolled in school.

Figure 5.1 Follow-Up Question Logic for Current Education



* A consistency check between the current and highest education was triggered if the respondent (R) (1) received a bachelor's degree but was (or will be) in the 4th year of college or lower, (2) completed the 12th grade without receiving a diploma but was (or will be) in the 12th grade or lower, or (3) received a general educational development certificate of high school completion but was (or will be) in the 12th grade or lower.

[Table 5.1](#) shows the names of the edited variables that corresponded to the questions in the current education section.

Table 5.1 Names for Current Education Variables

Question	Measure	Name
QD11 ¹	Additional edited variable for the highest level of education (in addition to edits from the consistency checks)	EDUHGHEDT
QD17, QD17a, QD17b	Currently going to school	EDUSCHLGO
QD18 ²	Current or upcoming grade in school	EDUSCHGRD
QD19 ²	Full-time or part-time student status	EDUFULPAR
QD20 ²	Number of days missed school in the past 30 days because of sickness or injury	EDUSCKMON
QD20DKRE ²	Estimated number of days missed school in the past 30 days because of sickness or injury	EDUSCKEST
QD20 QD20DKRE	Number of days missed school in the past 30 days because of sickness or injury (combined; see Section 5.2.6.6)	EDUSCKCOM
QD21 ²	Number of days missed school in the past 30 days because of skipping or “cutting” school	EDUSKPMON
QD21DKRE ²	Estimated number of days missed school in the past 30 days because of skipping or “cutting” school	EDUSKPEST
QD21 QD21DKRE	Number of days missed school in the past 30 days because of skipping or “cutting” school (combined; see Section 5.2.6.6)	EDUSKPCOM

¹ Question QD11 was in the front-end demographics section at the beginning of the interview but was the source variable for EDUHGHEDT.

² Questions QD18 through QD21DKRE applied to respondents who reported in QD17 through QD17b that they were currently in school or planned to return to school when their break was over. See [Figure 5.1](#) for additional information on routing within the current education section.

The remainder of this section discusses the following issues that were relevant to editing of the current education variables:

- overall routing logic for the current education section (Section 5.2.6.1);
- edits involving respondents’ school enrollment status (Section 5.2.6.2);
- general edits involving the last and current grade for respondents who were in school (Section 5.2.6.3);
- specific edits involving the last and current grade for respondents aged 12 to 17 who were in school (Section 5.2.6.4);
- specific edits involving the last and current grade for respondents aged 18 or older who were in school (Section 5.2.6.5); and
- miscellaneous edits involving the current education variables (Section 5.2.6.6).

[Table B.1](#) in Appendix B describes edits for the current education variables pertaining to current school enrollment, last and current grade, and associated variables if respondents were

currently in school. [Table B.1](#) also is intended as a companion to the discussion of the editing procedures for the current education variables in Sections 5.2.6.2 through 5.2.6.5.

5.2.6.1 Routing Logic for Current Education

As shown in [Figure 5.1](#), determining whether respondents were currently going to school was critical for determining whether they would be asked further questions about their education. However, respondents might answer the question about going to school (QD17) as “no” if they were not going to school because they were on holiday break or vacation from school, although the question instructed them not to. Therefore, certain respondents who did not report in question QD17 that they were currently going to school were asked follow-up questions (if they were aged 12 to 25 and their highest reported level of education from question QD11 was an associate’s degree or lower) to determine whether they were on a holiday or vacation break from school (question QD17a), and if so, whether they intended to return to school once their break was over (question QD17b).

If respondents reported in QD17 that they were not currently going to school (QD17 = 2) but they were on break from school (QD17a = 1) and they intended to return to school once their vacation or break was over (QD17b = 1), the edited variable EDUSCHLGO for going to school (based on data from QD17, QD17a, and QD17b) was set to a value of 1 (“yes”) to indicate that the respondents should be considered to be going to school. Otherwise, the response from QD17 was carried over to EDUSCHLGO. That assignment from QD17 included situations in which respondents reported in QD17a that they were not on vacation break from school, or if they reported in QD17b that they were on break but did not intend to return to school once their break was over.

If respondents were currently going to school or were on break but intended to return, the wording of subsequent questions varied according to their current status. For question QD18, respondents who reported in QD17 that they were currently going to school (QD17 = 1) were asked to report their current grade in school (i.e., at the time of the interview). For respondents who were on vacation break from school but intended to return to school once their break was over (QD17b = 1), question QD18 asked for the grade that they would be in once they returned from their vacation break. Similarly, for the question about full-time or part-time student status (question QD19) respondents who were currently going to school were asked “Are you a full-time student or a part-time student?” Those who were on break from school but intended to return to school were asked “Will you be a full-time or a part-time student?”

Based on the logic shown in [Figure 5.1](#), a key aspect of editing the current education variables involved assigning legitimate skip codes (Section 2.4.2) according to respondents’ eligibility to be asked questions in this section. For example, if respondents reported that they were part-time students (EDUFULPAR = 2), the edited variables corresponding to questions QD20 through QD21DKRE (EDUSCKMON, EDUSCKEST, EDUSKPMON, and EDUSKPEST) were assigned legitimate skip codes. Additional edits to assign legitimate skip codes to self-administered education variables are described in [Table B.1](#).

5.2.6.2 Edits Involving Current School Enrollment

Because the routing logic in the current education section was based on respondents' reports of whether they were going to school (or not), an important aspect of editing the variables in this section involved editing of the variable EDUSCHLGO for going to school. If respondents were not currently going to school (EDUSCHLGO = 2), then remaining education variables starting with EDUSCHGRD were assigned legitimate skip codes (Section 2.4.2).

5.2.6.3 General Edits Involving the Last and Current Grade

The current school grade from question QD18 could be inconsistent with the highest level of education that the respondent reported completing in question QD11 in the front-end demographics section. Consistency check questions QD18CC01 to QD18CC04 gave respondents the opportunity to resolve inconsistencies between QD11 and QD18, although the consistency checks did not cover all possible inconsistencies between these two questions. Therefore, edits were implemented to cover other inconsistencies in the data.

For respondents who had completed the 11th grade or lower, one might expect the current grade in QD18 to be one grade level higher than the response in QD11. In addition, no editing was done when the current grade reported in QD18 was the 11th grade or lower and was the same as the highest grade reported in QD11 because respondents could have been repeating a grade.

As noted in Section 4.2.9, editing of the front-end education variable EDUHIGHST (highest grade completed, corresponding to question QD11) with respect to data in QD18 for a respondents' current grade was limited to the consistency checks between QD11 and QD18 because QD18 was in a separate section of the interview. However, a second variable, EDUHGHEDT, was created as part of editing of the current education section in which the highest level of education underwent further edits. Consequently, the front-end education variable EDUHIGHST would not be affected by future changes that might occur in the content of current education variables. Yet, the EDUHGHEDT variable might in some situations be a more accurate reflection of the highest level of education that respondents had completed.

The general education edits discussed in the remainder of this section were invoked if an inconsistency occurred in the answers from QD11 and QD18 that had not triggered a consistency check between these two questions. An algorithm was used to handle these types of inconsistencies and to restrict the edits to data for respondents aged 12 to 17. This algorithm is discussed in detail below.

For respondents aged 12 to 17, a series of arrays was set up that mapped out the highest grade and current grade that would be *expected* relative to a respondent's current age, assuming an orderly progression from one grade level to the next highest level. [Table 5.2](#) shows a matrix that maps the current age with an acceptable range of grades in the United States.

Table 5.2 Mapping of Current Age with Acceptable Range of Grades

Current Age	12	13	14	15	16	17
Acceptable Completed Grades	4-8	5-9	6-10	7-11	8 through some college credit	9 through some college credit
Acceptable Current Grades	5-9	6-10	7-11	8-12	9 through college	10 through college

For example, one might expect most people in the United States to have completed the 6th grade by the time they are 12 years old. It would therefore not be unreasonable for someone to be aged 12 and to be currently in the 7th grade depending on when the respondent was interviewed. The algorithm was limited to those younger than 18 because response options in QD11 for levels beyond the 11th grade do not refer to single years of education. Level 12 in QD11 refers to the receipt of a high school diploma, and level 13 refers to completion of the 12th grade without the receipt of a diploma.

In addition, the algorithm allowed for some deviation relative to the expected ages, as shown in [Table 5.2](#). For example, if a respondent was aged 12 and was currently in the 9th grade, this was still considered to be an acceptable pattern because the respondent might have received accelerated education beyond the expected grade.

Most edits were done separately depending on whether respondents were aged 12 to 17 or 18 or older. The rationale for doing edits separately for these two age groups was that most respondents aged 12 to 17 would not be expected to have graduated from high school. In contrast, adults aged 18 or older could report in question QD11 that they had received a high school diploma, some education beyond high school but no degree, an associate's degree, a bachelor's degree, or a graduate or professional degree. [Table B.1](#) provides additional details on edits for EDUHGHEDT and EDUSCHGRD depending on whether respondents were aged 12 to 17 or 18 or older.

5.2.6.4 Edits Involving the Last and Current Grade for Respondents Aged 12 to 17

For respondents aged 12 to 17, the highest grade completed and the current grade were considered to be consistent with the respondent's age if what was reported was within 2 years of the grades shown in [Table 5.2](#). Thus, if a respondent was aged 12, the algorithm considered completion of any grades from the 4th through the 8th to be sufficiently consistent with the respondent's age. Similarly, for respondents aged 12, the algorithm considered current enrollment in any grades from the 5th through the 9th to be sufficiently consistent with the respondent's age.

Therefore, the following four data combinations were possible:

- both the completed grade and the current grade were considered to be consistent with the respondent's age;

- the highest completed grade was consistent with the respondent's age, but the current grade was not;
- the current grade was consistent with the respondent's age, but the highest completed grade was not; or
- neither the highest completed grade nor the current grade was consistent with the respondent's age.

Separate edits were done according to the four combinations of data patterns described immediately above. The following edits were done if both the completed grade and current grade were consistent with the respondent's age:

- If the current grade was more than two grade levels higher than the highest completed grade, the current grade was edited to be consistent with the highest grade because the latter was an interviewer-administered variable. For example, if a respondent was aged 12, reported completing the 6th grade, and reported currently being in the 9th grade (QD18 = 9), then the edits logically inferred that the respondent currently was in the 7th grade because this answer was more consistent with both the respondent's current age and the highest completed grade. The edited variable for current grade (EDUSCHGRD) was assigned a code of 27 (i.e., 7th grade LOGICALLY ASSIGNED).
- If the current grade was lower than the highest completed grade, the editing gave precedence to the reported grade that would yield the most consistent result relative to the respondent's age. In particular, if accepting the report of the highest grade and inferring that the respondent was currently in the next highest grade would yield a current grade that was inconsistent with the respondent's age, then the edited variable EDUHGHEDT for the highest level of education instead was assigned a value consistent with the current grade. Suppose, for example, that a 12-year-old respondent reported currently being in the 6th grade but having completed the 7th grade. Accepting the answer that the respondent was currently in the 6th grade and inferring that the respondent had completed the 5th grade would be more consistent with the respondents' current age than would be the converse (i.e., accepting that the respondent had completed the 7th grade and inferring that the respondent was currently in the 8th grade). In this example, EDUHGHEDT was assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If the highest completed grade was consistent with the respondent's age but the current grade was not, the highest completed grade was accepted by default. This assumption was made if the current grade was lower than the highest completed grade or the current grade was more than two grade levels higher than the highest completed grade. The edited current grade EDUSCHGRD was therefore assigned a value to indicate a current grade level that was 1 year higher than the highest completed grade. For example, if the respondent reported completing the 10th grade, EDUSCHGRD was assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).

If the current grade was consistent with the respondent's age but the highest completed grade was not, the editing procedures accepted the current grade by default. Thus, if a 12-year-old

respondent reported last completing the 3rd grade and reported currently being in the 6th grade, this edit would identify the current 6th grade as being consistent with an age of 12; that is, being in the 6th grade would be more consistent with the respondent's current age compared with completing the 3rd grade. In this example, EDUHGHEDT was assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If neither the current grade nor the reported highest grade was consistent with the respondent's age, the following was done:

- If the current grade was lower than the highest grade that was reported, the algorithm picked the answer that was closest to the expected grade, based on the matrix shown in [Table 5.2](#). The variable with the more inconsistent data was assigned a bad data code. This edit allowed for situations where respondents may have fallen behind where they would be expected to be grade-wise (e.g., if they had been held back a year).
- If the current grade was exactly two grade levels higher than the reported highest completed grade and the highest completed grade was higher than what would be expected for the respondent's age, no further editing was done. Otherwise, the created current variable EDUHGHEDT was assigned a bad data code. This edit was designed to allow for situations where a respondent might be on an accelerated track.
- If the current grade was more than two grade levels higher than the reported highest completed grade and it was lower than the expected current grade, then the value was retained for the current grade. The variable EDUHGHEDT was assigned a bad data code. In other situations, both EDUHGHEDT and EDUSCHGRD (i.e., the edited current grade) were assigned codes of bad data. The rationale for the first edit was that, if EDUSCHGRD was lower than the expected current grade, EDUSCHGRD would be more consistent with the expected current grade and the respondent's age than what the reported highest completed grade would be.

5.2.6.5 Edits Involving the Last and Current Grade for Respondents Aged 18 or Older

Editing of EDUHGHEDT and EDUSCHGRD for adult respondents aged 18 or older took into account levels 13 to 20 in question QD11 that indicate the receipt of a high school diploma, college degree, or partial completion of the requirements for a diploma or degree. Editing of EDUHGHEDT and EDUSCHGRD for adult respondents also took into account the respondents' current age. For example, if respondents aged 18 or 19 reported completing the 10th grade or lower and also reported that they were currently in the first or second year of college, then EDUHGHEDT was coded as bad data, because being in the first or second year of college was consistent with an age of 18 or 19.

5.2.6.6 Miscellaneous Edits Involving the Current Education Variables

Categorical follow-up questions (QD20DKRE and QD21DKRE) captured data from respondents who did not know or refused to report an exact number of school days (ranging from 0 to 30) that they missed in the past 30 days because they were sick or injured (QD20) or because they skipped school (QD21), respectively. These follow-up questions were intended to reduce the

amount of missing data for the number of days that respondents missed school for respondents who might find it easier or more acceptable to choose a category (i.e., a range) that best fit the number of days that they were absent. The follow-up questions included response choices for absences on the following number of days in the past 30 days: (1) 0 days, (2) 1 day, (3) 2 days, (4) 3 to 5 days, and (5) 6 or more days. A sixth response choice was available in QD20DKRE for respondents to indicate that school was not in session during the past 30 days. Respondents who chose the sixth response category were not asked questions regarding number of days of school missed because of skipping.

QD20DKRE (edited variable EDUSCKEST) and QD21DKRE (edited variable EDUSKPEST) were not used to edit EDUSCKMON (based on QD20) or EDUSKPMON (based on QD21). However, data from these questions were used to create composite variables (EDUSCKCOM and EDUSKPCOM) that combined information from these follow-up questions with the corresponding lead question (QD20 or QD21, respectively). These composite variables had a continuous range, similar to the edited variables EDUSCKMON and EDUSKPMON.

Edits for creating EDUSCKCOM and EDUSKPCOM are described below.

- If the lead variable (e.g., EDUSCKMON, corresponding to QD20) had a valid response from 0 to 30, then the corresponding composite variable (e.g., EDUSCKCOM) was assigned the value from the lead question.
- If the lead variable had a response of “don’t know” or “refused” and the corresponding follow-up variable (e.g., EDUSCKEST, corresponding to question QD20DKRE) indicated that respondents were absent for 0 days, 1 day, or 2 days, then the value from the follow-up question was assigned to the corresponding composite variable.
- If the lead variable had a response of “don’t know” or “refused” and respondents chose a range of days in which they were absent (i.e., 3 to 5 days or 6 or more days), then the midpoint of the range was assigned to the composite variable.
 - If respondents reported being absent on 3 to 5 days, then the corresponding composite variable (e.g., EDUSCKCOM) was assigned a value of 4 days.
 - If respondents reported being absent on 6 or more days, then the corresponding composite variable was assigned a value of 18 days, or the midpoint of 6 to 30 days.
- If EDUSCKMON had a response of “don’t know” or “refused” and EDUSCKEST indicated that school was not in session in the past 30 days, then EDUSKPMON, EDUSKPEST, and EDUSKPCOM were assigned legitimate skip codes (Section 2.4.2).
 - If QD21DKRE had been skipped because a value of 0 to 30 had been reported in QD21, then EDUSKPEST was coded as 99.

Respondents who reported in QD20DKRE that school was not in session (QD20DKRE = 6) were not asked questions QD21 and QD21DKRE about skipping school in the past 30 days. Therefore, if respondents reported in QD20DKRE that school was not in session, then EDUSKPMON, EDUSKPEST, and EDUSKPCOM were coded 99.

[Table B.1](#) discusses additional editing of variables in the current education section other than those pertaining to current enrollment and grade level. These included situations in which the data were consistent with respondents' status as going to school but other potential inconsistent responses were observed, such as responses for absences from school in the past 30 days.

5.2.7 Military Families

Questions QD10d and QD10e were used to identify whether respondents had immediate family members who were currently serving in the U.S. military (i.e., individuals serving in one of the branches of the U.S. Armed Forces, in a Reserve component, or in the National Guard).

Questions about family members who are currently serving in the U.S. military provide relevant information for supporting the well-being of military families.

In QD10d, a list of specific family members was presented to respondents, and respondents were asked to consider only those family members. Respondents then were asked whether any of these family members were currently serving in the U.S. military.

If respondents reported in QD10d that any of their family members were currently serving in the U.S. military, then they were asked in QD10e to select from the list of specific family members which family member or members were currently serving.

[Table 5.3](#) shows the names of the edited variables that corresponded to the questions for military family members.

Table 5.3 Names for Military Family Variables

Question	Measure	Name
QD10d	Any immediate family member currently serving in the U.S. military	MILTFAMILY
QD10e1	Spouse is currently serving in the U.S. military	MILTPOUS
QD10e2	Unmarried partner is currently serving in the U.S. military	MILTPARNR
QD10e3	Mother is currently serving in the U.S. military	MILTMOTHR
QD10e4	Father is currently serving in the U.S. military	MILTFATHR
QD10e5	Son or sons is/are currently serving in the U.S. military	MILTSON
QD10e6	Daughter or daughters is/are currently serving in the U.S. military	MILTDAUGH
QD10e7	Brother or brothers is/are currently serving in the U.S. military	MILTBROTH
QD10e8	Sister or sisters is/are currently serving in the U.S. military	MILTSISTR

As noted previously, if respondents answered QD10d as "yes," they were asked to report which members of their immediate family (as specified in QD10d) were currently serving in the U.S. military. Because QD10e was an "enter all that apply" question, separate variables were created for the family relationships in the question. Codes of 1 or 2 (Yes, or No not entered, respectively) were assigned to the edited variables corresponding to QD10e (see Section 2.4.4). If respondents answered QD10d as "no," then legitimate skip codes were assigned to MILTPOUS through MILOTHREL (Section 2.4.2).

Two additional edits were implemented for MILTMOTHR, MILTFATHR, MILTSON, and MILTDAUGH when respondents' ages were considered to be inconsistent with their reports of one of these family members currently serving in the U.S. military.

- If respondents were aged 50 or older and reported that their mothers or fathers were currently serving, then MILTMOTHR or MILTFATHR were assigned codes of 85 for bad data, as applicable.
- If respondents were aged 12 to 25 and reported that sons or daughters were currently serving, then MILTSON or MILTDAUGH were assigned codes of 85, as applicable.

5.2.8 Employment and Workplace

Respondents aged 15 or older were asked questions about their current employment, employment history, and characteristics of their workplace (if applicable). Question QD26 asked whether respondents worked in the week prior to the interview. If respondents reported that they did not work in the past week, they were asked in question QD27 whether they had a job or business. Respondents were then routed through different branches of work-related questions depending on how they answered these two key questions, as shown in [Figure 5.2](#). For example, respondents who worked in the past week were asked questions to determine full-time or part-time work status (i.e., whether they usually worked 35 or more hours per week), whether they ever had a period of unemployment in the past 12 months, the number of days they missed work in the past 30 days because they were sick or because they did not want to be at their workplace, and characteristics of their workplace, particularly with respect to alcohol and other drug policies at their workplace. Similarly, respondents who did not work in the past week and did not have a job were routed into questions relevant for people who currently were not working, such as why they did not have a job, whether they made specific efforts to find work in the past 30 days, and the month and year when they last worked for pay.

An important aspect of editing the work-related variables involved identification of situations where questions had been legitimately skipped (Section 2.4.2). A second key aspect of processing the work-related variables was to use the data to establish respondents' current work status ([Figure 5.2](#)). As noted previously, a single, recoded work status variable named WRKSTATWK was created that served as the starting point for creation of a simplified edited variable (EDWRKSTAT) and a final, statistically imputed employment status variable (IRWRKSTAT). WRKSTATWK was created from the following final variables: WRKDPSTWK (whether respondents worked in the past week), WRKHADJOB (whether respondents had a job if they did not work in the past week), WRK35WKUS (whether respondents usually worked 35 or more hours per week), WRKRSNOT (reason for not working in the past week despite having a job), WRKRSNJOB (reason for not having a job in the past week), WRKEFFORT (made specific efforts to find work in the past 30 days), and WRKDPSTYR (whether the respondents had a job in the past 12 months). Based on the data in these variables, respondents aged 15 or older were assigned to one of the categories in WRKSTATWK that are listed in [Table 5.4](#).

Figure 5.2 Routing Logic for Respondents in the Employment Status Section

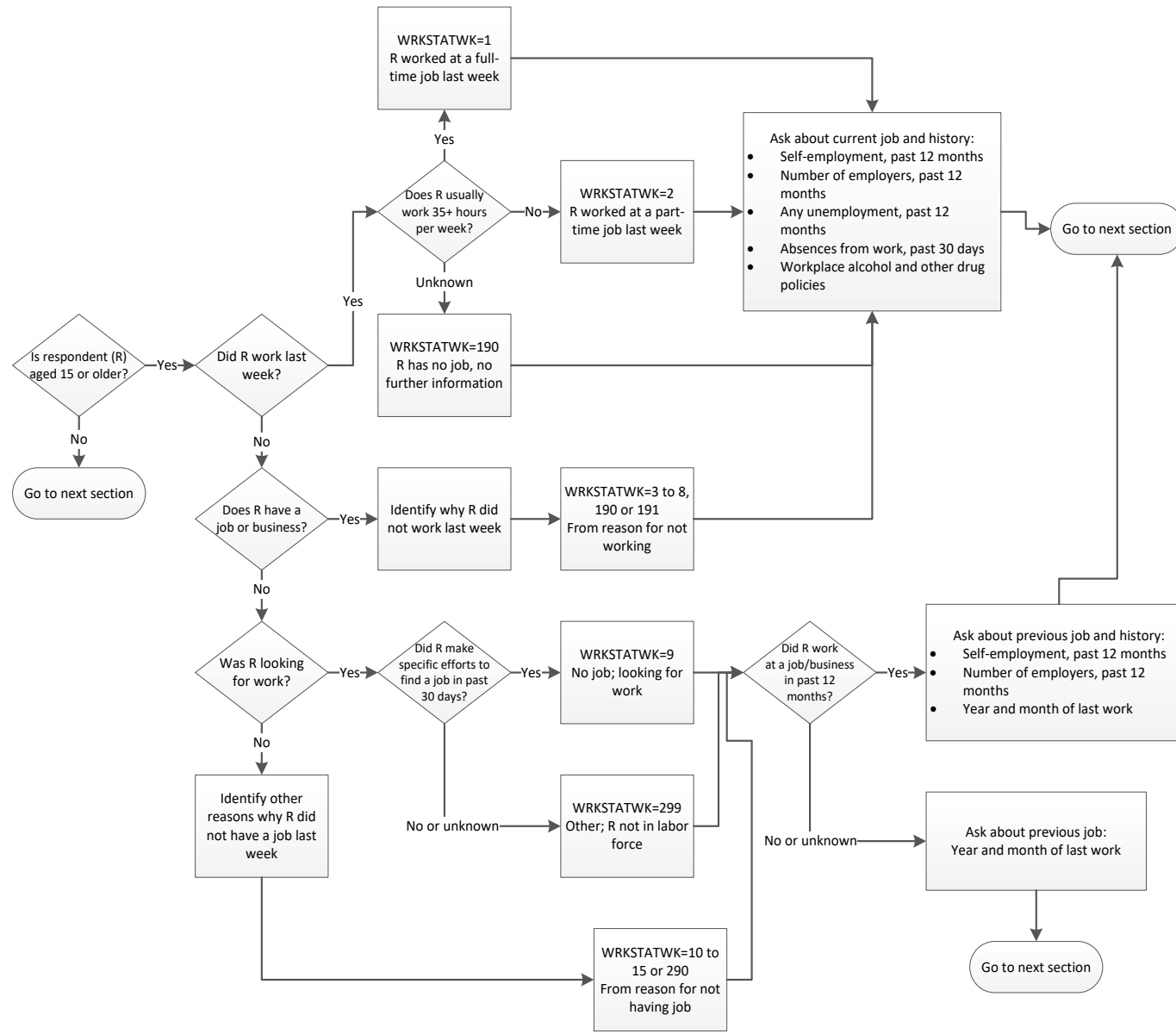


Table 5.4 Categories of WRKSTATWK

Code	Employment Situation	Code	Employment Situation
1	Worked at full-time job, past week	11	No job: keeping house full time
2	Worked at part-time job, past week	12	No job: in school/training
3	Has job but out: vacation/sick/temp absence	13	No job: retired
4	Has job but out: layoff, looking for work	14	No job: disabled for work
5	Has job but out: layoff, not looking for work	15	No job: didn't want a job
6	Has job but out: waiting to report to new job	190	Has full-time job, reason for not working unknown
7	Has job but out: self-employed, no business past week	191	Has part-time job, reason for not working unknown
8	Has job but out: in school/training	199	Has job, no further information
9	No job: looking for work	290	No job, no further information
10	No job: layoff, not looking for work	299	Other, not in labor force

NOTE: Additional codes for WRKSTATWK in the 900 series have their standard meanings in NSDUH: Don't know (994), Refused (997), Blank (998), and Legitimate skip (999).

In addition, respondents who reported in question QD31 that they did not have a job but were looking for work were not classified as being unemployed unless they reported in WRKEFFORT that they had made specific efforts in the past 30 days to find work (such as making contacts with someone about a job, sending out resumes or job applications, or placing or answering ads). If respondents reported that they did not have a job but were looking for work but WRKEFFORT was not answered as "yes," they were classified as not in the labor force (code 299) in WRKSTATWK.

If respondents did not know or refused to report whether they worked in the past week, WRKDPSTYR was checked for indications of whether respondents worked in the past year. Respondents who indicated in WRKDPSTYR that they did not work in the past 12 months were classified as not having a job (WRKSTATWK = 290). Otherwise, if respondents did not provide information on whether they worked in the past week (i.e., QD26 answered as "don't know" or "refused"), WRKSTATWK was assigned the corresponding code of "don't know" or "refused."

[Table B.2](#) describes edits for employment variables pertaining to whether respondents had a job in the past week, the number of days absent from work in the past 30 days, the number of weeks without a job in the past 12 months, the number of hours worked in the past week, the year and month that respondents last worked for pay, and self-employment. For example, if respondents reported in question QD39a that they never worked for pay, respondents were instructed to enter a response of 9991. When the month question QD39b had been skipped because a response of 9991 had been entered in QD39a, the edited month variable WRKLASTMO was assigned a code of 91. Documentation for 9991 (or 91) was as follows: 9991 = NEVER WORKED AT A JOB OR BUSINESS.

A final procedure that is discussed for the editing of the employment variables (not presented in [Table B.2](#)) concerns the creation of the base variable EDWRKSTAT, which was used to create

the imputation-revised employment status variable. EDWRKSTAT was derived from WRKSTATWK and WRK35WKUS. WRK35WKUS was used in some situations to determine whether employed respondents were employed full time or part time. Specifically, EDWRKSTAT was derived in the following manner:

EDWRKSTAT =

- 99, if the respondent is 12 to 14 years old; else
- 1 (full time), if WRKSTATWK = 1 or 190, or if WRKSTATWK = 3, 6, 7, 8, or 199 and WRK35WKUS = 1; else
- 2 (part time), if WRKSTATWK = 2 or 191, or if WRKSTATWK = 3, 6, 7, 8, or 199 and WRK35WKUS = 2; else
- 3 (unemployed), if WRKSTATWK = 4, 5, 9, or 10; else
- 4 (other), if WRKSTATWK = 11, 12, 13, 14, 15, 290, or 299; else
- 5 (part time or full time), if WRKSTATWK = 3, 6, 7, 8, or 199 and WRK35WKUS was missing (i.e., greater than 2); else
- missing.

5.3 Imputation for Back-End Demographics Variables

For the back-end demographics section, three imputation-revised variables were created. The first, IRWRKSTAT, was an employment status variable; the second, IRIMMBORNUS, was an indicator of whether the respondent was born in the United States; and the third, IRIMMENTAGE, recorded the age at which the respondent entered the United States. These imputation-revised variables were created from the base variables EDWRKSTAT, IMMBORNUS, and ENTRYAG2. ENTRYAG2 is the variable for the age at which an immigrant to the United States entered this country. It was derived from the respondent's age and the amount of time an immigrant lived in this country (IMMYRSNUM or IMMOSNUM). Employment status, the born in the United States indicator, and age at which the respondent entered the United States had item response rates above 95 percent.

These variables were processed in three separate, single-member imputation sets, using the single response propensity (RP)/single prediction (PRD) type of the predictive mean neighborhood (PMN) method described in Chapter 3. See Appendix D for details on the covariates used in the RP and PRD models for demographic variables.

5.3.1 Employment Status (Imputation Set 1)

The first self-administered demographic variable that underwent imputation was the employment status variable. The imputation process was straightforward except for one feature: the handling of the response EDWRKSTAT = 5, where it was known that the respondents were employed but it was not known whether they were employed part time or full time. These responses were handled in the hot-deck step in a separate missingness pattern, with a single logical constraint and a modified predictive mean vector. The final imputation-revised variable IRWRKSTAT had five levels: employed full time, employed part time, unemployed, other, and a skip code for respondents aged 12 to 14. Two age groups were used

for RP and PRD modeling: 15 to 25 and 26 or older. The hot-deck step was implemented separately for three age groups: 15 to 17, 18 to 25, and 26 or older.

5.3.1.1 Response Propensity Step

The response propensity models for imputation set 1 utilized the preliminary analysis weight, PANALWT. For the first RP model, the domain included all unit respondents aged 15 to 25. For the second RP model, the domain included all respondents aged 26 or older. In both situations, item respondents were those with EDWRKSTAT values of 1, 2, 3, or 4.

5.3.1.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the employment status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. For both age groups, the outcome variable had four levels, which mapped to the first four levels of EDWRKSTAT.

5.3.1.3 Hot-Deck Step

The predicted means from the prediction step played a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This allowed for parallel processing, which decreases the time required for implementation. Each item nonrespondent in the hot-deck step was assigned one of two missingness patterns. Item nonrespondents with a missing value for EDWRKSTAT were handled in the first missingness pattern, which used the full predictive mean vector and no logical constraints. Item nonrespondents with EDWRKSTAT = 5 were handled in the second missingness pattern, which applied a logical constraint to ensure that the donor was employed (either full time or part time). Also, conditional probabilities were used to take advantage of the partial information that was available. Instead of using the model's predicted probabilities directly, a single predicted mean was derived using a conditional probability, which was the probability that the respondent was employed full time, given that the respondent was employed. In addition to the segment and delta likeness constraints, a third likeness constraint (i.e., donor's age must be within 5 years of recipient's age) was applied in the hot-deck step. See [Table E.14](#) in Appendix E for more details on missingness patterns and constraints for employment status.

5.3.1.4 Recodes for Additional Analyses

IRWRKSTAT18 was a direct recode of IRWRKSTAT and AGE. For respondents who were younger than 15 or older than 17, IRWRKSTAT18 and IRWRKSTAT were equivalent. For 15- to 17-year-olds, responses for IRWRKSTAT were overwritten with a code indicating that respondents were too young to have their employment status recorded for the variable. This was the same code that was used for 12- to 14-year-olds for IRWRKSTAT (and IRWRKSTAT18).

5.3.2 Immigrant Status, Born-in-U.S. Indicator (Imputation Set 2)

The second self-administered demographic variable that underwent imputation was the born-in-U.S. variable, IMMBORNUS. This was a dichotomous variable with very few missing responses. The RP, PRD, and hot-deck steps were run separately within three age groups: 12 to 17, 18 to 25, and 26 or older. The imputation procedure was straightforward and is described in the next three sections.

5.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for IMMBORNUS.

5.3.2.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the born-in-U.S. indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.⁶⁹ The outcome variable was IMMBORNUS. The single predicted mean was the predicted probability that the respondent was born in the United States.

5.3.2.3 Hot-Deck Step

The hot-deck step for the born-in-U.S. indicator included a single predicted mean from the prediction step, no logical constraints, and only the segment and delta likeness constraints. IMMBORNUS was the base variable for imputation and the imputation-revised version was called IRIMMBORNUS. Details on the hot-deck step are available in [Table E.16](#).

5.3.3 Immigrant Status, Age of Entry (Imputation Set 3)

The age of entry variable was created only for respondents who were not born in the United States as defined by IRIMMBORNUS. This yielded a small domain. As a result, imputations were not conducted separately within age groups. The models were likely to be improved when age groups were combined because (1) none of the response categories were sparsely populated, leading to more robust models and (2) sufficiently large donor pools were ensured in the hot-deck step. Details on the procedures applied to the age of entry variable are explained in the next three sections.

5.3.3.1 Response Propensity Step

The response propensity model for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all respondents who were not born in the United States as defined by IRIMMBORNUS. Item respondents were those domain members with a nonmissing value for ENTRYAG2.

⁶⁹ In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

5.3.3.2 Prediction Step

Using the adjusted weights that were outputs of the RP step, the predicted mean for an immigrant's age of entry was estimated using a linear regression model, as implemented by the REGRESS procedure in SUDAAN. To control the upper and lower bounds of predicted means for age of entry, it was necessary to perform a logit transformation on the response variable. The response variable in the model was the immigrant age at entry as a proportion of the continuous version of current age, IMMCONTAGE, which is calculated from the interview date (INTDATE) and date of birth (BRTHDATE). The expression of the proportion of a respondent's life spent outside of the United States is $P_i = Y_i/N_i$, where i indexes the respondents, $Y_i = \text{Age at Entry}_i$, and $N_i = \text{Continuous Age}_i$ (IMMCONTAGE).

After the weight adjustment, the following empirical logit transformation was used as the response variable in a weighted linear univariate regression:

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

This transformation was nearly equivalent to the standard logit transformation,

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

which was not used because this transformation is unstable for respondents who entered the country at their current age (such that $P_i = 1$).

5.3.3.3 Hot-Deck Step

Two logical constraints were utilized in the hot-deck step for the age of entry variable. Both involved the respondent's age. One required that the donor's age of entry be less than or equal to the recipient's current age. The other required that the difference between the recipient's current age and the donor's age of entry be less than 1 year if the recipient lived in the United States for less than 1 year (as indicated by QD16a) or greater than 1 year if the recipient lived in the United States for more than 1 year. The only base variable was ENTRYAG2 and its imputation-revised counterpart was IRIMMENTAGE. The segment and delta likeness constraints were applied in the hot-deck step. Details on the hot-deck step are available in [Tables E.17](#) through [E.19](#).

6. Editing and Imputation for the NSDUH Substance Use Variables

6.1 Introduction

This chapter provides documentation of procedures for the editing and statistical imputation of variables in the 2023 National Survey on Drug Use and Health (NSDUH) that were used to estimate substance use in the lifetime, past year, and past month periods for NSDUH tables and reports.⁷⁰ In addition, the editing and imputation of related substance use variables, such as frequency of use, age and date of substance initiation, reasons for using, and ways of obtaining prescription psychotherapeutics, are discussed. As a prerequisite for reviewing this chapter, readers are encouraged to familiarize themselves with the content and terminology in Chapters 1 through 3 of this report because these three chapters provide an overview to the NSDUH interview and the general approach to editing and imputation of NSDUH data, with information on specific processes that are common across the survey in Chapters 2 and 3.

For all substance use sections (tobacco, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription stimulants, and prescription sedatives), variables corresponding to questions for key substance use measures first underwent editing to (1) replace missing values with codes that indicated that the questions did not apply, (2) replace missing values with nonmissing values that could logically be inferred from other data, (3) make logical inferences when answers to related questions were inconsistent, and (4) identify inconsistent or inexact responses to be resolved through imputation. Except where indicated, edited variables for these measures then underwent imputation to replace missing values with nonmissing values or to replace inexact responses (e.g., use at some point in a respondent's lifetime but no definite period for most recent use) with specific ones.

6.1.1 Edited (but Not Imputed) Substance Use Variables

Some substance use variables underwent editing but no further imputation. Many of these variables pertain to prescription psychotherapeutic drugs⁷¹ or tobacco products. In particular, the following groups of prescription psychotherapeutic drug variables were edited, but missing values were retained rather than being replaced through statistical imputation:

- any use or misuse (if applicable) in the past 12 months of specific prescription drugs;⁷²
- the age at first misuse (AFM), year of first misuse (YFM), and month of first misuse (MFM) for specific prescription drugs that were misused in the past 12 months;

⁷⁰ The past year refers to the period within 12 months of the interview date (i.e., past 12 months), and the past month refers to the period within 30 days of the interview date (i.e., past 30 days). Past month use also is referred to as "current" use in NSDUH reports.

⁷¹ Prescription psychotherapeutic drugs include pain relievers, tranquilizers, stimulants, and sedatives.

⁷² The exception is that any use and misuse in the past 12 months were imputed for OxyContin®, prescription fentanyl, and benzodiazepines. See Section 6.5.2 for more information.

- misuse of any prescription psychotherapeutic drug within an overall category (e.g., pain relievers) within the past 30 days when the respondent was drinking alcohol or within a couple of hours of when the respondent drank alcohol;
- ways in which respondents misused prescription psychotherapeutic drugs within an overall category in the past 12 months (e.g., without a prescription of the respondent's own);
- the specific prescription drug within a given category that respondents last misused in the past 12 months;
- use of a needle to inject prescription stimulants in the past 12 months or past 30 days;
- reasons for misusing the last prescription drug within a given category in the past 12 months; and
- how respondents (or a friend or relative, if applicable) obtained the last prescription drug within a given category that they misused in the past 12 months.

In addition to the above, Section 6.2.7 discusses editing procedures for tobacco variables that did not undergo further imputation. Also, the variable corresponding to the usual number of drinks of alcoholic beverages that respondents consumed on days when they used alcohol in the past 30 days (corresponding to question AL07) is used in editing other variables that subsequently undergo imputation (e.g., frequency of consumption of four or more drinks for females or five or more drinks for males on a single occasion in the past 30 days, also referred to as the frequency of "binge" alcohol use). However, the variable corresponding to question AL07 itself does not undergo imputation. The editing procedures in place for the variable corresponding to question AL07 are discussed in Section 6.2.4.7.

6.1.2 "Parent" and "Child" Substance Use Measures

In this chapter, some of the substance use measures refer to a general category (e.g., any hallucinogen use), and other measures refer to one or more subcategories within a general category (e.g., LSD, PCP, or Ecstasy as types of hallucinogens). These categories are described using the terms "parent" for the general category and "child" for the subcategory. Parent/child pairs often (but not always) occurred in sections that included multiple gate questions.⁷³ For example, the hallucinogens section includes questions about lifetime use of specific hallucinogens to determine the respondent's overall lifetime use or nonuse of any hallucinogen. Included in these questions are specific gate questions for LSD (LS01A); PCP (LS01B); Ecstasy, also known as MDMA (LS01F); ketamine (LS01I); the tryptamine hallucinogens DMT, AMT, or "Foxy" (LS01J); and *Salvia divinorum* (LS01K).⁷⁴ However, parent/child pairs also could appear in separate sections, such as for any use of cocaine and specific use of crack cocaine (i.e., if respondents reported lifetime use of any form of cocaine). Not every substance that falls into a particular category is considered a child substance (e.g., specific tranquilizer substances are not

⁷³ See Section 2.2.3 for a discussion of the gate question terminology.

⁷⁴ The hallucinogens section includes questions about additional specific hallucinogens (e.g., peyote in LS01C). However, these are not considered "child" substances because follow-up information is not collected about initiation and recency.

considered child substances of the overall tranquilizer category). [Table 6.1](#) shows the measures with parent/child relationships and the data that were collected for them.

Table 6.1 Substance Use Measures with Parent/Child Relationships

Parent Measure	Child Measure	Estimates from Parent Data	Estimates from Child Data	"Other" Lifetime Use Indicator ¹
Any Cigarette Use	Daily Cigarette Use	Initiation, recency, 30-day frequency	Initiation	No
Any Alcohol Use	Past Month Binge Alcohol Use	Initiation, recency, 12-month frequency, 30-day frequency	30-day frequency ²	Not applicable
Any Cocaine Use	Crack Use	Initiation, recency, 12-month frequency, 30-day frequency	Initiation, recency, 12-month frequency, 30-day frequency	No
Any Hallucinogen Use	Use of LSD, PCP, Ecstasy, Ketamine, DMT/AMT/"Foxy," or Salvia divinorum	Initiation, recency, 12-month frequency, 30-day frequency	Initiation (LSD, PCP, and Ecstasy only), recency (all child hallucinogens)	Yes

¹ See Section 6.4.1.

² The consumption of alcohol section includes questions about the initiation of binge alcohol use (Section 8.5.5). However, in keeping with the general principle of not editing across sections of the interview (Section 2.4.1), data in the alcohol section were not edited for consistency with data in the later consumption of alcohol section. Moreover, variables in the consumption of alcohol section (including variables for the initiation of binge alcohol use) were not statistically imputed.

6.2 Editing Procedures for the Substance Use Variables Other Than Prescription Drugs

Consistent with prior years, the substance use measures collected for tobacco, nicotine vaping, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, and methamphetamine included (but were not limited to) the following:⁷⁵

- lifetime use (or nonuse);
- initiation of use (i.e., age at first use [AFU], and for recent initiates, the month of first use [MFU] and year of first use [YFU]);
- most recent use;
- frequency of use in the past 12 months (for nonprescription substances other than tobacco products and nicotine vaping); and
- frequency of use in the past 30 days (for all nonprescription substances except for pipe tobacco).

This section discusses procedures for editing the substance use variables for tobacco (i.e., cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol,

⁷⁵ Questions regarding lifetime use and most recent use of nicotine vaping were added to the emerging issues section of the 2021 NSDUH. Beginning in 2022, these questions were moved to the tobacco section, and additional questions were asked regarding initiation of use and frequency of use in the past 30 days.

marijuana, cocaine in any form, crack cocaine, heroin, hallucinogens, inhalants, and methamphetamine.

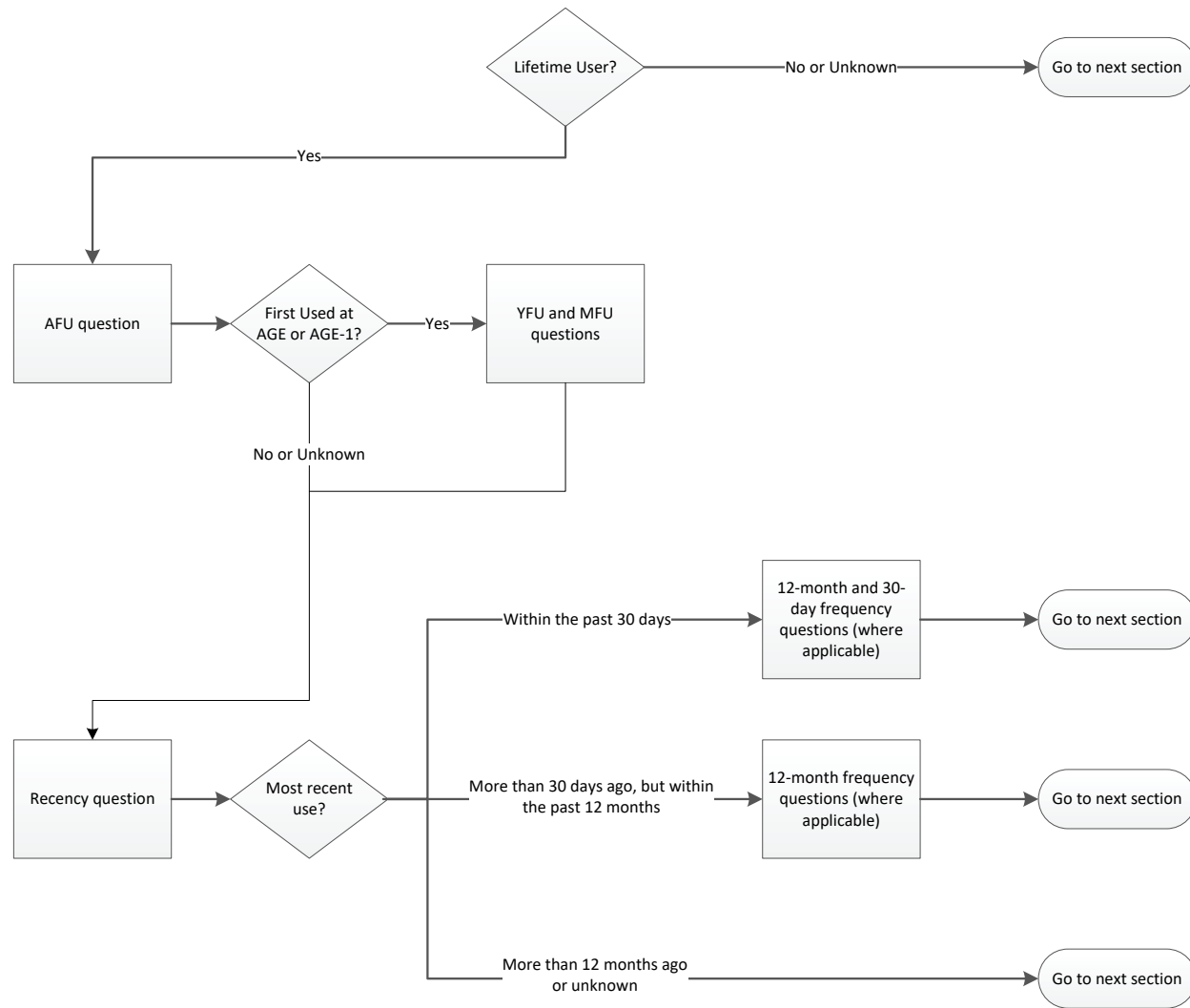
[Figure 6.1](#) provides an overview of the general routing logic for the substance use sections of the interview other than for prescription drugs.⁷⁶ Edits for the substance use variables in these sections encompassed the following key activities:

- edits of the lifetime use variables (i.e., gate questions), where respondents indicated whether they have ever used the substance of interest (Section 6.2.1);
- edits of the recency-of-use variables, where respondents who indicated lifetime use of the substance indicated when they last used it (Section 6.2.2);
- edits involving users of only “child” drugs for hallucinogens (i.e., where use of all other specific hallucinogens was not reported) (Section 6.2.3);
- edits of the 12-month and 30-day frequency variables, where respondents who indicated use of a substance in the 12 months or 30 days prior to the interview indicated the number of days they used that substance in the period of interest (Section 6.2.4);
- edits for variables pertaining to the initiation of use, where respondents who indicated lifetime use of a substance indicated when they first used it (Section 6.2.5);
- edits for lifetime daily cigarette use (Section 6.2.6);
- edits for tobacco use variables that do not undergo imputation (Section 6.2.7); and
- edits that were applied as part of the imputation processing (Section 6.4).

In connection with each of these edits, the discussion in the remainder of Section 6.2 focuses on relevant issues or inconsistencies in the data that needed to be addressed through logical editing. As noted in Section 1.2, the skip logic in the computer-assisted interviewing (CAI) instrument limited the chances for respondents to be routed to questions where they could give answers that were inconsistent with their answers to previous questions. For example, if respondents reported never using marijuana in question MJ01, the CAI program did not ask additional questions that would presume use of marijuana at least once. Nonetheless, the CAI program did not completely eliminate all opportunities for respondents to provide inconsistent answers, though most processing of the CAI data was relatively straightforward, and the issues discussed in this section were not widespread relative to the total number of respondents.

⁷⁶ For additional information about the content and logic in the substance use section of the questionnaire, see the computer-assisted interviewing specifications on the [NSDUH Data Collection](#) web page.

Figure 6.1 Routing Logic for Respondents in the Substance Use Sections Other Than for Prescription Drugs



AFU = age at first use; MFU = month of first use; YFU = year of first use.

6.2.1 Edits of Lifetime Use Variables

As discussed in connection with the usable interview criteria (Section 2.2.3), respondents were asked whether they had ever used the following:

- tobacco products (cigarettes, smokeless tobacco, cigars, or pipe tobacco);
- nicotine vaping
- alcohol;
- marijuana;
- cocaine (and crack cocaine, if they had ever used cocaine in any form);
- heroin;
- hallucinogens;
- inhalants; and
- methamphetamine.

For hallucinogens and inhalants, respondents were asked a series of questions to establish whether they had ever used one or more specific types of drugs within that category (e.g., LSD, PCP, peyote). Only those respondents who indicated lifetime use of that substance (or lifetime use of one or more specific hallucinogens or inhalants) were asked more detailed questions about that substance, or any hallucinogen or inhalant (including situations in which respondents initially refused to answer a question about their lifetime use of a substance but then changed their answer to “yes” on follow-up; see Section 2.2.3).

For hallucinogens and inhalants, however, there was no overall question about lifetime use or nonuse of any drug in these categories. Rather, lifetime use of any drug in the overall category for these two sections could be deduced from one or more affirmative answers regarding lifetime use of any of the specific drugs in that category. Because the sections for hallucinogens and inhalants included a question about the lifetime use of “any other” drug in that category, answers of “no” for lifetime use of all of the specific drugs in that category (i.e., including lifetime use of “any other”) logically indicated that respondents never used any drugs in that category.⁷⁷

Processing of the gate variables for substances other than prescription drugs (tobacco, nicotine vaping, alcohol, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine) established whether (1) respondents had used a substance of interest at least once or, for hallucinogens or inhalants, any substance in that category; (2) they had never used the substance (or any substance in that category); or (3) lifetime use or nonuse of the substance could not be determined. In addition to answering these gate questions as “yes” or “no,” respondents could answer them as “don’t know” or “refused.” As noted in Section 2.4.2, final responses of “don’t know” and “refused” were treated by the CAI skip logic as equivalent to situations where respondents never used the substance of interest. For hallucinogens and inhalants, the CAI skip logic treated combinations of responses of “no,” “don’t know,” and “refused” to the questions about individual drugs in the absence of any affirmative response to these questions in the same manner as if respondents had answered all of these questions negatively. In these situations where a gate question was answered as “don’t know” or “refused,” respondents’ lifetime use or nonuse status was treated as unknown because these responses did not provide conclusive information one way or the other.⁷⁸ Respondents with unknown lifetime use/nonuse status were subsequently imputed to be lifetime users or nonusers, as described in Section 6.5.

This step of the editing procedures also involved assignment of “bad data” codes to lifetime substance use variables (i.e., equivalent to missing data) if potential patterned responses previously had been identified (Sections 2.3.2 and 2.4.3). For the substance use sections pertaining to smokeless tobacco, cigars, nicotine vaping, alcohol, marijuana, cocaine, crack,

⁷⁷ Respondents who refused to answer all questions about the lifetime use of specific inhalants were asked a follow-up question (INREF) for whether they ever used any type of inhalant. See Section 6.2.1.2 for details on the use of INREF for classifying respondents as lifetime users or nonusers of inhalants.

⁷⁸ For multiple gate drugs, this pattern included situations where respondents answered one or more of the gate questions as “don’t know” or “refused” but did not answer any of the other gate questions affirmatively.

heroin, and methamphetamine,⁷⁹ these edits involved assignment of codes for bad data for the lifetime substance use question (e.g., question AL01 for alcohol). For hallucinogens and inhalants, these edits involved assignment of codes for bad data to all lifetime variables in that section (e.g., data in questions LS01A through LS01K for hallucinogens, including question LS01H for any other hallucinogen and associated “OTHER, Specify” data for hallucinogens).

6.2.1.1 Creation of Lifetime Use Variables for Single-Gate Substances

The creation of edited variables for lifetime use or nonuse of substances for which there was a single-gate question for lifetime use or nonuse⁸⁰ was typically straightforward because there were overall questions about lifetime use or nonuse. In most instances, codes corresponding to respondents’ original answers simply were assigned to the relevant variable (e.g., 1 = Yes; 2 = No). When respondents were routed to a follow-up question because they initially refused to answer the first question about lifetime use or nonuse (Section 2.4.2), their response to this follow-up question was taken as the final response for the lifetime use variable.

One exception to these procedures involved lifetime use or nonuse of crack cocaine. Specifically, the response to question CC01, regarding lifetime use of any form of cocaine, governed whether respondents were asked questions about crack cocaine or were skipped out of questions for crack cocaine. Therefore, the lifetime crack use variable CRKEVER (corresponding to question CK01) was assigned a code of 91 (i.e., never used cocaine) when COCEVER (corresponding to CC01) was answered as “no.” If respondents reported lifetime use of cocaine but answered question CK01 as “no,” then CRKEVER was assigned a code of 2.

6.2.1.2 Creation of Lifetime Use Variables for Hallucinogens and Inhalants

In the hallucinogens and inhalants sections, if respondents answered all gate questions that indicated lifetime use or nonuse of specific drugs as “no,” the edit procedures assigned a code of 91 to the entire series of variables (e.g., LSD; PCP; peyote; mescaline; psilocybin; Ecstasy; ketamine; DMT, AMT, or Foxy; *Salvia divinorum*; and any other hallucinogen).⁸¹ This edit was done to indicate that the respondents had never used any of the drugs in that category. By contrast, a code of 2 (i.e., “no”) in edited variables for lifetime use of specific drugs within a broader category had the following meaning:

- respondents were users of at least one drug in the category but had never used the specific drug of interest (e.g., if respondents were lifetime users of LSD but not PCP); or

⁷⁹ Because of the requirement of the usable interview criteria that respondents had to have defined data for lifetime use or nonuse of cigarettes (Section 2.2.3), assignment of the missing value of bad data for lifetime use of cigarettes would result in an interview being reclassified as not usable.

⁸⁰ Substances with single-gate questions for lifetime use were tobacco products (cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, and methamphetamine.

⁸¹ In the hallucinogens section, this included situations in which respondents initially refused to report whether they had ever used LSD, PCP, or Ecstasy but then changed their answer(s) to “no” on follow-up.

- respondents reported never using the particular drug of interest but answered other questions in the series as “don’t know” or “refused” (e.g., if respondents did not know whether they had ever used LSD but knew they had never used PCP).

Variables also were created to indicate whether respondents had ever used one or more drugs within the overall drug categories. These variables were called HALLUCEVR (ever used any hallucinogen) and INHALEVER (ever used any inhalant). These summary variables could take on the following values:

- **1 (“Yes”)**: The summary variable for a given category was coded as 1 (i.e., “yes”) if respondents answered “yes” for lifetime use of at least one specific drug in the overall category. The summary variable also was coded as 1 if respondents initially refused to answer all lifetime questions within one of these sections and then reported on follow-up that they had used some drug in that category at least once. For example, if at least one affirmative answer appeared in questions about lifetime use of inhalants (including the follow-up question INREF), INHALEVER was coded as 1.
- **91 (“Never Used”)**: This code of 91 could be assigned to the overall summary variable for one of two reasons: (1) if respondents answered all questions about lifetime use of individual drugs within the category as “no” or (2) if respondents were routed to a follow-up question because they refused to answer *all* questions about lifetime use of individual drugs within a category and then reported that they never used that type of drug. As noted previously, this code of 91 also was assigned to all of the individual lifetime use variables within that section. For example, if respondents initially refused to answer all questions about lifetime use of specific inhalants, they were asked whether they had ever used any type of inhalant. If respondents answered this follow-up question as “no,” the summary variable INHALEVER was assigned a code of 91, and all individual lifetime inhalant variables were assigned a code of 91 as well.
- **94 (“Don’t know”)**: This edit applied only to the summary variable INHALEVER for inhalants. If respondents were routed to INREF because they refused all of the gate questions for inhalants and they answered INREF as “don’t know,” then INHALEVER was assigned a final code of 94.
- **97 (“Refused”)**: The summary variable was assigned a code of 97 if respondents initially refused to answer *all* questions about lifetime use for hallucinogens or inhalants and then continued to refuse when asked the follow-up questions about lifetime use. In this instance, the edited lifetime use variables for the individual drugs within that section also retained a code of 97. For example, INHALEVER was assigned a code of 97 if a respondent refused to answer questions IN01A through IN01L for inhalants and also refused to answer the follow-up question INREF.

- **98 ("Blank"):** The summary variable was assigned a code of 98 if respondents had one or more answers of "don't know" or "refused" to questions about lifetime use of specific drugs within the category (i.e., other than the previous situation in which respondents refused all questions) and they had no reports of lifetime use of any of these drugs. In this situation, the code of 98 indicated that lifetime use or nonuse for the overall category was unknown, even if one or more questions about lifetime use of specific drugs was answered as "no." For example, INHALEVER was assigned a code of 98 if some questions in IN01A through IN01L were answered as "no" but other questions were answered as "don't know" or "refused." For hallucinogens, respondents were also assigned a code of 98 in the summary variable HALLUCEVR if they answered all of the lifetime questions in LS01A through LS01H as "don't know."

6.2.1.3 Editing of "OTHER, Specify" Drug Variables for Nonusers of Hallucinogens or Inhalants

If respondents had never used any of the drugs in the series of multiple gate questions for hallucinogens or inhalants, they would not have been routed to questions where they could specify the use of some other drug in that overall category. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with a code of 9991 to indicate that the questions were skipped because the respondents had never used that class of drugs.

Similarly, if respondents reported in the preceding multiple gate questions for hallucinogens or inhalants that they used at least one drug in the category, but they never used some other drug besides the ones they were asked about, they were legitimately skipped out of the questions that asked them to specify what "other" drug they had used. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with a code of 9999 (LEGITIMATE SKIP).

Functionally, the codes of 9991 and 9999 both indicate that respondents legitimately skipped out of the "OTHER, Specify" questions in that drug's section. However, the 9991 code provides for analysts the extra level of detail that the respondents were legitimately skipped out of these questions because they had never used anything within that category of drugs.

6.2.1.4 Use of "OTHER, Specify" Drug Data to Edit Lifetime Use Variables for Hallucinogens and Inhalants

When respondents specified the use of other hallucinogens or inhalants, these "OTHER, Specify" data were used to edit related variables *within the same section* (Sections 2.4.1 and 2.4.5). For example, if respondents did not report lifetime use of a specific drug in the initial question(s) (e.g., PCP in LS01B or LSREF2), but they specified using it as another drug in the category within the same interview section (e.g., "marijuana laced with PCP" as another hallucinogen), the respondents were logically inferred to be lifetime users of that drug (e.g., PCP for "OTHER, Specify" data within the hallucinogens section). In such circumstances, the code 3 = Yes LOGICALLY ASSIGNED was assigned to the edited lifetime use variable for that drug.

This edit was especially relevant to the following hallucinogens: LSD; PCP; Ecstasy; ketamine; the tryptamines DMT, AMT, or “Foxy”; and *Salvia divinorum*. For these hallucinogens, lifetime use or nonuse was imputed if respondents had missing data for the lifetime use questions. Consequently, no imputation would be needed for these drugs if respondents did not know or refused to report whether they had used these drugs when asked directly but the “OTHER, Specify” data in that section indicated use.

With few exceptions, “OTHER, Specify” data that referred to street names or slang terms were not used in editing the lifetime variables for hallucinogens or inhalants, because the meaning of a particular slang term could vary by region, could apply to more than one drug, or could change over time. Use of street or slang terms to infer lifetime use of a specific hallucinogen or inhalant was generally restricted to situations where that term was supplied to respondents as a synonym for that drug. For example, “angel dust” was listed in question LS01B as an alternative term for PCP. Thus, situations where respondents specified “angel dust” as some other hallucinogen were used to infer lifetime use of PCP, if question LS01B had not already been answered as “yes.” Similarly, respondents who specified “Molly” (with no other information) as some other hallucinogen but who did not report lifetime Ecstasy use have been inferred to be Ecstasy users; however, reports of “Molly” that could be associated with another drug (e.g., the stimulant slang terms “White Molly” or “Black Molly”) have not been used to infer Ecstasy use. Other slang terms that were classified with particular hallucinogens included “Shrooms” (coded as Psilocybin/Mushrooms), “X,” and “X-TC” (the latter two coded as Ecstasy).

In addition, as noted in Section 2.4.1, “OTHER, Specify” data for hallucinogens were not used to edit variables for inhalants, or vice versa. For example, if respondents specified an inhalant (e.g., gasoline) as some other “hallucinogen” that they had ever used, then this report was not used in editing the data for lifetime inhalation of gasoline or lighter fluid from the inhalants section. However, the “OTHER, Specify” data for hallucinogens in this example would indicate to analysts when respondents specified the use of a drug that fits another category. This approach also assumed that some respondents may specify the use of other drugs according to their functional properties. For example, the definition of hallucinogens (i.e., drugs that “often cause people to see or experience things that are not real”) could result in respondents thinking of other drugs besides hallucinogens that alter one’s perception of reality.

6.2.2 Edits of Recency-of-Use Variables for Substances Other Than Prescription Drugs

Edits of the variables that establish when respondents last used a substance of interest are generally the most critical.⁸² The recency-of-use variables for substances other than prescription

⁸² The recency-of-use variables in this section for substances other than prescription drugs were based on direct questions that asked respondents to report the last time that they used a particular substance (e.g., “How long has it been since you last used marijuana or hashish?”).

drugs⁸³ were the precursors for the final measures that established the prevalence of use in the past 30 days, past 12 months, and lifetime.⁸⁴

The skip logic in the CAI instrument for substances other than prescription drugs limited the kinds of information that were available for use in editing the recency-of-use variables. Respondents who answered a gate question (or all multiple gate questions) for a given substance as “no” (i.e., never used that substance) were not given the opportunity to answer additional questions as though they were users of that substance. Similarly, respondents who reported that they last used a substance “more than 12 months ago” were not given the opportunity to answer further questions in that section about use in the past 12 months or past 30 days, as though they were more recent users than what they had originally indicated ([Figure 6.1](#)).

6.2.2.1 Edits to Recency Variables for Nonusers

As noted in Sections 2.2.3 and 2.4.2, respondents were skipped out of all remaining questions about use of a particular substance when they answered “no” to a given single-gate question⁸⁵ or when they answered “no” to all of the individual gate questions for hallucinogens and inhalants.

A code of 91 (i.e., “never used”) typically was assigned to the recency variables when it was determined unambiguously that respondents had never used the substance of interest. For crack cocaine, recency variables were also assigned a code of 91 if the lifetime variable CRKEVER was coded as 91 because the respondent reported never using cocaine in any form. Similarly, for hallucinogens and inhalants, the recency variables were assigned a code of 91 if all of the lifetime use variables corresponding to the gate questions had been assigned a code of 91 because respondents had never used any of the individual drugs (Section 6.2.1.2).

6.2.2.2 Recency Periods in the Questionnaire

In the sections for alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine, the following standard codes for recency applied, based on the available responses to the question(s) about most recent use:

- 1 = Within the past 30 days,
- 2 = More than 30 days ago but within the past 12 months, or
- 3 = More than 12 months ago.

⁸³ These substances were tobacco products (cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine.

⁸⁴ For example, the final recoded variable ALCFLAG for published estimates of the lifetime use of alcohol was created from the imputed recency variable IRALCRC. IRALCRC was created from the edited recency variable ALCREC. If IRALCRC indicated use in any period, then ALCFLAG was coded as 1 (i.e., ever used). If IRALCRC indicated that respondents never used alcohol, then ALCFLAG was coded as 0 (i.e., never used).

⁸⁵ Substances with single-gate questions for lifetime use were tobacco products (cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, and methamphetamine.

In the tobacco sections for cigarettes, smokeless tobacco, cigars, and nicotine vaping, lifetime users first were asked whether they used the relevant tobacco product in the past 30 days (e.g., question CG05 for cigarettes). If respondents reported that they did not use the particular tobacco product in that period, they were asked to report most recent use prior to the past 30 days (e.g., question CG06 for cigarettes). Recency questions for these tobacco products and nicotine vaping included categories for most recent use (1) more than 30 days ago but within the past 12 months, (2) more than 12 months ago but within the past 3 years, and (3) more than 3 years ago.⁸⁶

For these tobacco products and nicotine vaping, single edited recency variables were created that combined the data for the questions about use in the past 30 days and most recent use in later periods, if applicable. For these substances, the following standard codes for recency applied, based on the available responses to the question(s) about most recent use:

- 1 = Within the past 30 days,
- 2 = More than 30 days ago but within the past 12 months,
- 3 = More than 12 months ago but within the past 3 years, or
- 4 = More than 3 years ago.

The CAI instrument included follow-up probes for respondents who were lifetime users of a given substance but did not know or refused to report when they last used it.⁸⁷ Respondents who initially did not know when they last used a substance were asked to give their “best guess” of when they last used it. Respondents who initially refused to report when they last used a substance were asked to reconsider answering the question. If respondents changed their initial answer of “don’t know” or “refused” to report a definite period when they last used the substance of interest, then that information served as the starting point for subsequent editing of the substance’s recency variable. The standard recency codes described previously in this section were assigned according to the specific period of most recent use that was reported in these follow-up probes. In the absence of any inconsistencies between the recency-of-use answers in the follow-up probes and other data within a given substance’s section, these answers from the follow-up probes were accepted as final and were incorporated within the edited recency variable.

6.2.2.3 Logically Assigned Recency Periods among Lifetime Users

The following codes could be assigned to the edited recency variables when respondents reported lifetime or more recent use:

- 8 = Used at some point in the past 12 months LOGICALLY ASSIGNED,
- 9 = Used at some point in the lifetime LOGICALLY ASSIGNED,

⁸⁶ A recency variable was not created for pipe tobacco because lifetime users of pipe tobacco were asked only if they smoked tobacco in a pipe in the past 30 days.

⁸⁷ For cigarettes, smokeless tobacco, cigars, and nicotine vaping, these follow-up probes were asked if respondents reported that they did not use a given tobacco product or nicotine vaping in the past 30 days but they answered “don’t know” or “refused” to the follow-up question about most recent use more than 30 days ago.

- 11 = Used in the past 30 days LOGICALLY ASSIGNED,
- 12 = Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED,
- 13 = Used more than 12 months ago LOGICALLY ASSIGNED,⁸⁸
- 14 = Used more than 12 months ago LOGICALLY ASSIGNED,⁸⁹
- 19 = Used more than 30 days ago LOGICALLY ASSIGNED,
- 29 = Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED, and
- 39 = Used at some point within the past 3 years LOGICALLY ASSIGNED.

In the edited recency variables for substances other than prescription drugs,⁹⁰ for example, respondents were assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) if they reported lifetime use of a substance but continued on follow-up to answer “don’t know” or “refused” regarding when they last used it. These respondents were eligible to be statistically imputed to be users in any period, including in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago. Similarly, for respondents whose recency-of-use questions for LSD, PCP, Ecstasy, ketamine, DMT/AMT/“Foxy,” or *Salvia divinorum* had been skipped but they were logically inferred to be lifetime users of these drugs based on their responses to “OTHER, Specify” items (Section 6.2.1.4), their edited recency variables at least initially were assigned a code of 9.

In the recency variables for cigarettes, smokeless tobacco, cigars, and nicotine vaping, respondents were assigned a code of 19 (Used more than 30 days ago LOGICALLY ASSIGNED) if they reported that they did not use the tobacco product or nicotine vaping in the past 30 days, but they had missing data for the period more than 30 days ago when they last used it. These respondents were eligible to be statistically imputed to have most recently used a given tobacco product or nicotine vaping more than 30 days ago but within the past 12 months, more than 12 months ago but within the past 3 years, or more than 3 years ago.

Section 6.2.2.5 discusses additional situations in which codes were assigned to parent and child recency variables based on the editing procedures. In particular, codes of 11 or 12 could be assigned to a parent recency variable (e.g., any hallucinogen) based on data from a child recency variable (e.g., LSD). Recency variables did not require further statistical imputation when codes of 11 or 12 were assigned.

⁸⁸ A code of 13 applied to the recency variables for marijuana and hallucinogens. A code of 14 applied to the recency variables for cigarettes, smokeless tobacco, cigars, and nicotine vaping; this code included situations where it could be logically inferred that a respondent used more than 12 months ago but within the past 3 years or used more than 3 years ago.

⁸⁹ Codes of 14, 19, 29, and 39 applied only to the recency variables for cigarettes, smokeless tobacco, cigars, and nicotine vaping.

⁹⁰ These substances were tobacco products (cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine.

Recency variables that were assigned codes of 8, 9, 14, 19, 29, or 39 underwent further statistical imputation to assign a specific period of most recent use. These codes were used to place constraints during the subsequent imputation process for assigning the period of most recent use. For example, assigning a code of 8 to an edited recency variable constrained the imputation for the final recency to be within the past 30 days or more than 30 days ago but within the past 12 months. Similarly, for tobacco recency variables, a code of 14 constrained imputation to be more than 12 months ago but within the past 3 years or more than 3 years ago.

6.2.2.4 Application of “Flag and Impute” Procedures to Editing of Recency-of-Use Variables for Substances Other Than Prescription Drugs

The procedures used to edit the CAI recency-of-use variables for substances other than prescription drugs were referred to as the “flag and impute” procedures (Section 2.4.7). Under these procedures, the limited situations where potential inconsistencies existed between a respondent’s answer to a substance’s recency question and other data in that section were identified and flagged.

[Table 6.2](#) lists the usual types of inconsistencies that could occur between a substance’s recency variable and other variables in that section, and how these inconsistencies were handled through the flag and impute procedures. In the situations described in [Table 6.2](#), these inconsistencies were handled by statistically imputing final values for the affected recency variable and for the other variable(s) where the data were inconsistent with the respondent’s original answer to the recency question. Most of the codes that were described in Section 6.2.2.3 were assigned to the edited recency variables as a result of these edits. For example, if respondents reported first use of marijuana at their current age but also reported most recent use more than 12 months ago, both answers logically cannot be true. In this situation, the edited recency variable MJREC was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) and the inconsistent initiation variables were assigned codes for bad data (Sections 2.3 and 2.4.3).

Prior to implementing these flag and impute rules, initial edits checked for situations where respondents’ original answers to an AFU question might have been inconsistent with their recency of use but the revised AFU was not. For example, if respondents reported first using a substance at their current age, they were asked to verify this AFU. If the respondents reported that this AFU was not correct but then on follow-up did not know at what age they first used, or refused to answer, the edits updated the AFU to reflect this “don’t know” or refusal response. The rationale for this edit was that the respondents indicated that the initial answer was not correct. A final answer of “don’t know” or “refused” to an AFU question would not necessarily be inconsistent with a reported recency of use more than 12 months ago.

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables

Type of Inconsistency	Edits Implemented
Recency originally indicated use in the past 30 days, but use on 0 days in the past 30 days was confirmed (suggesting less recent use).	The edited recency variable was assigned a code of 8 (i.e., Used at some point in the past 12 months LOGICALLY ASSIGNED) to indicate that the respondent (R) was at least a user in the past 12 months and potentially a user in the past 30 days (Section 6.2.2.3). The 30-day frequency was set to missing (i.e., bad data).
Recency originally indicated use more than 30 days ago but within the past 12 months, but the 12-month frequency indicated use on more than 335 days in that period (suggesting past month use).	The edited recency variable was assigned a code of 8. The 12-month frequency data were set to missing.
Recency did not indicate use in the past 30 days, but the R reported first using the substance (or smoking cigarettes daily) in the same month as the interview took place (suggesting past month use).	If the recency originally indicated use more than 30 days ago but within the past 12 months, it was assigned a code of 8. If the recency originally indicated use more than 12 months ago (or was missing), it was assigned a code of 9 (i.e., Used at some point in the lifetime LOGICALLY ASSIGNED) to indicate that the R was at least a lifetime user (and potentially a user in the past 12 months or past 30 days). The values in the month of first use (MFU) and year of first use (YFU) that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes. The MFU and YFU variables also were set to bad data if the recency was missing and had been assigned a code of 9 because the initiation data would have suggested use in the past 30 days.
Recency did not indicate use in the past 30 days, but the R had other data suggesting initiation of use in the past 30 days (e.g., based on a comparison of the 12-month frequency and the maximum number of days that the R could have used the substance).	The edited recency variable was assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) to infer that the R was a past month user. No further editing was done to the variables indicating use in the past 30 days. This edit was an exception to the general “flag and impute” procedures.
Recency did not indicate use in the past 12 months, but the age at first use (AFU; or age at first daily use of cigarettes) equaled the R’s current age (suggesting past year use).	For alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine, the edited recency variable was assigned a code of 9. For tobacco products and nicotine vaping, if the lead 30-day question (e.g., CG05 for cigarettes) indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19 (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). Otherwise, the relevant tobacco or nicotine vaping recency variable was assigned a code of 9. The AFU that triggered the inconsistency with the recency-of-use answer was set to bad data. If the MFU and YFU were answered (i.e., not blank), the values in these variables were overwritten with bad data codes. (Month and year data were considered to be linked with the AFU data. Therefore, if the AFU was questionable, then the month and year also were considered to be questionable.)

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables (continued)

Type of Inconsistency	Edits Implemented
Recency did not indicate use in the past 12 months, but the R reported first using in a month and year that falls within 12 months of the interview date (including data for the month and year when the R reported first smoking cigarettes daily).	For alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine, the edited recency variable was assigned a code of 9. For tobacco products and nicotine vaping, if the lead 30-day question (e.g., CG05 for cigarettes) indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19 (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). Otherwise, the relevant tobacco or nicotine vaping recency variable was assigned a code of 9. The values in the MFU and YFU that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes.
For tobacco products and nicotine vaping, recency indicated use more than 3 years ago, but the AFU (or first daily use of cigarettes) indicated that the first use was within 2 years of the R's current age (suggesting use within the past 3 years). (Edits checking for indications of use at the R's current age were given priority over the condition described here because the former response pattern suggested use in the past year.)	The recency was assigned a code of 14 (Used more than 12 months ago LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 12 months ago (but potentially in the past 3 years). The AFU that triggered the inconsistency with the recency-of-use answer was set to bad data. If the MFU and YFU were answered (i.e., not blank), the values in these variables were overwritten with bad data codes.
For tobacco products and nicotine vaping, Rs did not use in the past 30 days but did not report a specific period of most recent use more than 30 days ago. However, Rs reported first use within 1 or 2 years of their current age (i.e., but not at the R's current age).	The recency was assigned a code of 29 (Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 30 days ago but within the past 3 years (but potentially within the past 12 months).
For tobacco products and nicotine vaping, Rs did not indicate whether they used in the past 30 days. However, Rs reported first use within 1 or 2 years of their current age (i.e., but not at the R's current age).	The recency was assigned a code of 39 (Used at some point in the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point within the past 3 years (but potentially within the past 30 days or past 12 months).

CAI = computer-assisted interviewing.

NOTE: Indications of most recent use included answers from follow-up probes for the recency questions.

6.2.2.5 Edits to Parent and Child Recency Variables

Important exceptions to the general flag and impute principles involved situations where inconsistencies existed between parent and child recency variables (any cocaine and crack cocaine; any hallucinogen use and LSD, PCP, Ecstasy, ketamine, DMT/AMT/"Foxy," or *Salvia divinorum* use). These are presented in [Table 6.3](#), along with a description of how the data were edited when specific types of inconsistencies occurred between related recency variables. In these special situations, indications of use of the child drug (e.g., crack cocaine) that were more recent than that indicated for the parent drug category (e.g., cocaine in any form) were used to logically infer more recent use of the parent drug category. For example, not all respondents

might make the connection that crack cocaine fits within the broader category of cocaine in general.

Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables

Child Recency Reported by Respondent or Assigned through Previous Editing Steps ¹	Parent Recency Reported by Respondent or Assigned through Previous Editing Steps ²	Final Edited Child Recency ¹	Final Edited Parent Recency ²
(1) Indicated use in past month.	Indicated use that is less recent than the past month.	Retained the recency reported by the respondent (R).	Logically inferred the R to be a past month user. Assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED).
(2) Indicated use more than 30 days ago but within the past 12 months.	Coded as 8, indicating use at some point in the past 12 months (Section 6.2.2.3).	Retained the recency reported by the R.	Retained the code of 8 to indicate that the R has used at some point in the past 12 months.
	Indicated use more than 12 months ago.	Retained the recency reported by the R.	Logically inferred the R to have last used more than 30 days ago but within the past 12 months. Assigned a code of 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).
	Coded as 9, indicating use at some point in the lifetime (Section 6.2.2.3).	Retained the recency reported by the R.	Logically inferred the R to be at least a past year user. Assigned a code of 8.
(3) Indicated use more than 12 months ago.	Coded as 8, indicating use at some point in the past 12 months.	Retained the recency reported by the R.	Retains the code of 8 to indicate that the R is at least a past year user.
	Coded as 9, indicating use at some point in the lifetime.	Retained the recency reported by the R.	Retained the code of 9 to indicate that the R is at least a lifetime user.
(4) Coded as 8, indicating use at some point in the past 12 months (see Table 6.2).	Indicated use more than 30 days ago but within the past 12 months.	Retained the code of 8 to indicate use at some point in the past 12 months.	Assigned a code of 8 to indicate use at some point in the past 12 months.
	Coded as 8, indicating use at some point in the past 12 months.	Retained the code of 8 to indicate use at some point in the past 12 months.	Retained the code of 8 to indicate use at some point in the past 12 months.
	Coded as 9, indicating use at some point in the lifetime.	Retained the code of 8 to indicate use at some point in the past 12 months.	Assigned a code of 8 to indicate use at some point in the past 12 months.

**Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies
between Parent and Child Recency Variables (continued)**

Child Recency Reported by Respondent or Assigned through Previous Editing Steps ¹	Parent Recency Reported by Respondent or Assigned through Previous Editing Steps ²	Final Edited Child Recency ¹	Final Edited Parent Recency ²
(5) Coded as 9, indicating use at some point in the lifetime.	Indicated use more than 30 days ago but within the past 12 months.	Retained the code of 9 to indicate use at some point in the lifetime.	Assigned a code of 8 to indicate use at some point in the past 12 months.
	Indicated use more than 12 months ago.	Retained the code of 9 to indicate use at some point in the lifetime.	Assigned a code of 9 to indicate use at some point in the lifetime.
	Coded as 9, indicating use at some point in the lifetime.	Retained the code of 9 to indicate use at some point in the lifetime.	Retained the code of 9 to indicate use at some point in the lifetime.

NOTE: These edits took place after inconsistencies had been identified between a recency variable and nonrecency variable (e.g., between the recency and the age at first use). For hallucinogens and the child drugs in that section, these edits also took place after the respondent had revised one or more answers in response to a consistency check. Further, for hallucinogens and the child drugs in that section, any inconsistencies that remain between a given recency variable and other nonrecency variables following inconsistency resolution were transferred back into the recency variables prior to implementation of these edits. For example, if the original answer to the hallucinogen recency disagreed with the age at first use but the revised recency in response to the consistency check did not, then the recency would be updated to reflect the revised value. Prior to implementation of the edits shown in the table, if the revised recency still disagreed with the age at first use, then the recency would be edited further to reflect the fact that the previous inconsistency still remained.

¹ Child drugs include crack, LSD, PCP, Ecstasy, ketamine, DMT/AMT/"Foxy," and *Salvia divinorum*.

² Parent drugs include any cocaine and any hallucinogens.

Therefore, if respondents reported last using any cocaine more than 30 days ago and also reported last using crack cocaine in the past 30 days, the edit procedures assigned a code of 11 to the edited cocaine recency variable COCREC (Section 6.2.2.3). This edit indicated logical inference that these respondents had used cocaine in any form in the past 30 days.

In addition, special patterns could remain in the data for cocaine and hallucinogens after most inconsistencies had been addressed and some related data elements were missing. Specifically, respondents could indicate that they first used any cocaine or any hallucinogen within 12 months of the interview date (e.g., first use at their current age) and indicate that they last used that drug more than 30 days ago but within the past 12 months (e.g., HALLUCREC = 2 for any hallucinogen use). Logically, then, if respondents who had used a child drug within a given category (e.g., LSD, PCP, Ecstasy, ketamine, DMT/AMT/"Foxy," or *Salvia divinorum*) had missing data on first use for one or more of their child drugs, it followed not only that they had to have first used a child drug at some point in the past 12 months, but also that they had to have last used the child drug at some point within the past 12 months. In this situation, both the parent recency (e.g., HALLUCREC) and child recency variable(s) (e.g., LSDREC, PCPREC, ECSREC) were set to values of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) so that they would be imputed consistently. If the parent recency indicated past

month use (e.g., HALLUCREC = 1) when initiation of the parent occurred in the past 12 months but a child recency (e.g., LSDREC) did not indicate use in the past 12 months, only the child recency was set to a value of 8.

Similarly, respondents could indicate that they first used any cocaine, or any hallucinogen, in the same month that they were interviewed and indicate that they last used the drug in the past 30 days but have missing data on first use for the child drug(s) within a category. In this situation, respondents who had used a child drug within that category (e.g., LSD) also were inferred to have last used that drug in the past 30 days. The edits assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) to the child recency variables (e.g., LSDREC = 11).

6.2.3 Edits Involving Users of Only Child Drugs for Hallucinogens

In addition to the situations described in [Tables 6.2](#) and [6.3](#), special edits were applied in situations for hallucinogens in which respondents reported lifetime use of only child drugs. The following specific patterns indicated that respondents were users of only the child drugs:

- one or more gate questions for the child drugs (i.e., LSD, PCP, Ecstasy, ketamine, DMT/AMT/"Foxy," or *Salvia divinorum*) were answered as "yes" and all other gate questions for hallucinogens were answered as "no"; or
- only the gate question for use of any other hallucinogen was answered as "yes" (i.e., including the gate questions for the child drugs) but the only report in the "OTHER, Specify" data for that section was for use of a child drug.⁹¹

Respondents were not considered to be users of only child drugs if they reported lifetime use of only child drugs but had responses of "don't know" or "refused" for any of the other gate questions. For example, suppose a respondent answered the question about lifetime use of LSD as "yes" and answered the question for lifetime use of peyote as "don't know" or "refused." If the respondent had no reports of lifetime use of hallucinogens from the other gate questions, they could have been lifetime users of peyote, which is not a child drug for hallucinogens.

6.2.3.1 Edits to Child Hallucinogen Variables for Users of Only One Child Drug

Users of only one child hallucinogen who answered the gate question for one of the child drugs as "yes" and answered all remaining gate questions as "no" were asked questions about initiation and recency for the overall parent category. The corresponding questions were

⁹¹ Respondents were considered to be users of only the child drug if they had codes only for "blank" or "bad data" in their "OTHER, Specify" data (see Section 2.3) in addition to the report for a child drug. Respondents were considered not to be users of only the child drug if they had codes for "don't know" or "refused" in their "OTHER, Specify" data in addition to the report for the child drug. These responses of "don't know" or "refused" were considered to be potential indications of use of another drug besides the child drug.

skipped for the child drug.⁹² Corresponding patterns occurred in the data if respondents reported lifetime use only of “other” hallucinogens and specified lifetime use of only a single child drug.

If respondents reported use of only one child hallucinogen, the values from the parent variables were assigned to the corresponding variables that had been skipped for the child drug. For example, if a respondent reported lifetime use of only LSD, the value from the edited AFU variable for hallucinogens (HALLUCAGE) and the value from the edited hallucinogen recency variable (HALLUCREC) were assigned to the corresponding variables for LSD (LSDAGE and LSDREC, respectively). Similarly, any values for the MFU and YFU for recent initiates of any hallucinogen (HALLUCYFU and HALLUCMFU, respectively) were assigned to the corresponding year and month variables for LSD. These values also were retained in the edited variables for use of any hallucinogen. Section 6.2.5 provides further details about the content of the questions in the substance use sections for initiation of use.

6.2.3.2 Edits for Lifetime Users of Multiple Child Drugs (but Only Child Drugs)

A set of edits were implemented for respondents who reported lifetime use of more than one of the six child hallucinogens (but only child drugs). These respondents were asked the questions for any hallucinogen and for the child drugs.

Some of these respondents reported more recent use of any hallucinogen than they reported for the individual child drugs. This pattern would be inconsistent for respondents who were lifetime users of only child drugs, but it would logically be plausible if the respondents were lifetime users of other hallucinogens in addition to the child drugs. In this situation, therefore, the response of 2 (i.e., no) in the variable for the lifetime use of any other hallucinogen (HALLUCOTH, corresponding to question LS01H) was replaced with a code of 5. This code had the following meaning: 5 = Yes (specific hallucinogen unknown) LOGICALLY ASSIGNED. That is, it was inferred that these respondents logically were lifetime users of some other hallucinogen, but the specific other hallucinogen was unknown. The corresponding “OTHER, Specify” variables for hallucinogens were assigned a code of 9998 (i.e., blank). Thus, the edit conserved information about more recent use of any hallucinogen rather than infer less recent use for consistency with the most recent use of the “child” hallucinogens.

Similar edits are described in Section 6.2.5 according to respondents’ answers to the initiation questions for any hallucinogen and for the child drugs LSD, PCP, and Ecstasy. Edits based on the initiation data did not apply to ketamine, DMT/AMT/“Foxy,” and *Salvia divinorum* because these child drugs did not include initiation questions.

⁹² Because the questionnaire included initiation and recency questions for LSD, PCP, and Ecstasy, the corresponding questions were skipped if respondents reported lifetime use of only one of these child drugs. Because the questionnaire included only recency questions for the child drugs ketamine, DMT/AMT/“Foxy,” or *Salvia divinorum*, the corresponding recency questions were skipped if respondents reported lifetime use of only one of these child drugs.

6.2.4 Edits of Frequency-of-Use Variables

Sections of the interview for substances other than prescription drugs⁹³ included questions about the number of days that respondents used different substances in the past 30 days or past 12 months (or the average number of days per week or days per month that they used in the past 12 months). These are referred to in this section as 30-day and 12-month frequency variables, respectively. Data from these frequency questions can be used to distinguish between occasional and more frequent users of a substance. For example, frequent users of alcohol and illicit drugs may represent a group who are potentially in need of substance use treatment or other services for their substance use. Similarly, regular users of tobacco products, such as people who smoked cigarettes every day in the past 30 days, probably represent a group that would have greater difficulty stopping their use of tobacco. In addition, the alcohol section included a question about the number of days that respondents consumed four or more drinks (for females) or five or more drinks (for males) per occasion in the past 30 days, a question which is used to construct measures of binge and heavy alcohol use in that period.⁹⁴

6.2.4.1 Preferred Ways of Reporting 12-Month Frequency

For the 12-month frequency determinations, respondents first were asked how they preferred to report their frequency of use in the past 12 months. Respondents could indicate a preference to report their frequency of use in one of three ways: (1) use on an average number of days per week in the past 12 months, (2) use on an average number of days per month in the past 12 months, or (3) the total number of days they used in the past 12 months. In particular, respondents who used a substance regularly in the past 12 months might find it easier to report their frequency of use in one of the first two ways as opposed to figuring the total number of days they used in that entire period. Conversely, respondents who used on only a few days in the past 12 months might prefer the third reporting method.

Individual variables (subsequently referred to as “source variables”) were created for the preferred way of reporting the 12-month frequency and for the associated follow-up questions (i.e., average number of days per week, average number of days per month, or total number of days in the past 12 months). An overall measure of the number of days that respondents used a substance in the past 12 months was created from these individual source variables.

For respondents who chose to report a total number of days that they used a substance, the 12-month frequency was the actual number of days that the respondent reported using the substance in the past 12 months (assuming no inconsistency with the 30-day frequency; these inconsistencies are discussed in Section 6.2.4.4). For respondents who chose to report an average use in days per week or days per month, the overall number of days that they used in the past 12 months was a calculated value. Specifically, answers in terms of the average

⁹³ These substances were tobacco products (cigarettes, smokeless tobacco, cigars, and pipe tobacco), nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine.

⁹⁴ Binge alcohol use refers to the consumption of four or more drinks for females or five or more drinks for males on a single occasion on at least 1 day in the past 30 days. Heavy alcohol use refers to binge alcohol use on 5 or more days in that period.

number of days used per week in the past 12 months were multiplied by 52, and answers in terms of the average number of days used per month in the past 12 months were multiplied by 12 in order to yield a calculated 12-month frequency. Although these latter two response options were averaged responses over the entire past 12 months (but respondents who last used a substance more than 30 days ago but within the past 12 months did not use in the month before the interview), no further adjustments were made to the calculated 12-month frequency value when respondents used the substance more than 30 days ago but within the past 12 months and they did not initiate use at some point in the past 12 months.

6.2.4.2 Editing of 12-Month and 30-Day Frequency Variables for Nonusers or Less Recent Users

If the lifetime gate question(s) and edited recency-of-use variable indicated that respondents had never used the substance of interest,⁹⁵ then edits at this step assigned a code of 91 to the 30-day frequency variable (where applicable)⁹⁶ and a code of 991 to the final 12-month frequency variable (where applicable). For questions on substances where respondents were asked to report their frequency of use in the past 12 months, a code of 91 (or 991) was assigned to the source variables pertaining to the preferred method of reporting the 12-month frequency (i.e., average number of days per week, average number of days per month, or total number of days used in the past 12 months), the average number of days per week, the average number of days per month, and total number of days used in the past 12 months.

Similarly, if the edited recency of use indicated that the respondent had used the substance but not in the period of interest, edits at this step assigned a code of 93 to the 30-day frequency variable and a code of 93 (or 993) to the 12-month frequency variable and related source variables that were used to create the 12-month frequency.

6.2.4.3 Editing of 12-Month and 30-Day Frequency Variables When the Recency Was Indefinite

If respondents were potentially users in the period of interest (i.e., there was some question about when respondents last used the substance) and the CAI program had skipped the 30-day or 12-month frequency questions, then the skipped variables retained a blank code. For example, if respondents reported lifetime use of a substance but they did not indicate a specific period for most recent use (Section 6.2.2.3), the frequency-of-use questions retained missing values. If the recency was imputed to indicate use in the past 30 days or past 12 months, the imputation procedures also replaced the missing data in these frequency-of-use variables with nonmissing values (Section 6.5).

Also, for alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine, if respondents reported last using a substance more than 12 months ago,

⁹⁵ For hallucinogens and inhalants, this set of responses meant that the respondent had never used any of the drugs in that category.

⁹⁶ For alcohol, this edit also applied to other 30-day variables, including the variable on the number of days in the past 30 days that respondents had four or more drinks (for females) or five or more drinks (for males) on a single occasion.

the CAI program skipped the questions pertaining to the frequency of use in the past 12 months and past 30 days. However, reporting first use of the substance at the respondents' current age would be inconsistent with the reported recency. As discussed in Section 6.2.2.4 and [Table 6.2](#), the flag and impute edit rules assigned a value to the recency variable to indicate that these respondents were users at some point in their lifetime, which could include use in the past 30 days or past 12 months. In this situation, the 12-month and 30-day frequency variables that had been skipped retained a blank value in case subsequent imputation might assign the respondents to a more recent category.

In addition, as discussed in Section 2.4.3, if respondents refused the lifetime gate question(s) and were skipped out of the 12-month and 30-day frequency questions (where applicable),⁹⁷ the edits at this step assigned refusal codes to the skipped frequency questions (i.e., the refusal was propagated). However, if respondents were skipped out of the 12-month and 30-day frequency questions because they answered the lifetime gate question(s) as "don't know," the edits retained codes of "blank" in the frequency variables.

6.2.4.4 Editing of the Frequency-of-Use Variables Based on Consistency Checks

Sections that contained both 12-month and 30-day frequency variables⁹⁸ included consistency checks between these variables. A consistency check was triggered in situations where the number of days that respondents reported using the substance in the past 30 days exceeded the number of days that the respondent used in the past 12 months.⁹⁹

If respondents did not know how many days they used a substance in the past 30 days or refused to give an answer, they were asked to give their best estimate of the number of days that they used. Respondents could estimate their 30-day frequency by choosing the category most likely to contain the number of days they used the substance: 1 or 2 days, 3 to 5 days, 6 to 9 days, 10 to 19 days, 20 to 29 days, or all 30 days. A consistency check also was triggered if the number of days that respondents reported using a substance in the past 12 months was lower than the minimum value for the number of days that respondents estimated using that substance in the past 30 days. For example, it would be inconsistent for respondents to report using marijuana on 6 to 9 days in the past 30 days and also to report using it on fewer than 6 days in the past 12 months.

If respondents revised either the 12-month or 30-day frequency data (or both) to make them consistent (i.e., such that the 12-month frequency was greater than or equal to the 30-day frequency following any updates done by the respondent), data from the consistency checks

⁹⁷ For the tobacco, nicotine vaping, alcohol, marijuana, cocaine, crack, and heroin variables, such a situation would occur if respondents initially refused the gate question and then refused again on follow-up.

⁹⁸ These sections were alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine.

⁹⁹ If respondents reported use on an average number of days per month or days per week in the past 12 months, the consistency check was triggered if the number of days respondents used in the past 30 days exceeded the calculated number of days respondents used in the past 12 months.

were taken as final. This edit included situations in which respondents resolved inconsistencies between their 12-month frequency and the minimum value for their estimated frequency of use in the past 30 days. If the 30-day frequency (or the minimum value for an estimated 30-day frequency) still was greater than the computed 12-month frequency despite a consistency check having been triggered, then the 12-month frequency was assigned a bad data code.

These edits based on data from consistency checks also applied to data that respondents entered in the follow-up questions for the 12-month and 30-day frequencies for hallucinogens. For example, when respondents were asked follow-up questions about their 30-day frequency of use (Section 6.2.4.6), consistency checks existed between the 12-month hallucinogen frequency and answers to the follow-up questions for the hallucinogen 30-day frequency. Thus, if respondents were routed to the follow-up questions for the hallucinogen 30-day frequency and the resulting 30-day frequency continued to be greater than the 12-month frequency after respondents were prompted to resolve the inconsistency, the hallucinogen 12-month frequency was assigned a bad data code.

In addition, a consistency check was triggered in the alcohol section if respondents reported that they binge drank (i.e., four or more drinks for females or five or more drinks for males in a single occasion) on more days in the past 30 days than they reported for the number of days in which they drank any alcohol in that period. If respondents made their overall 30-day frequency of alcohol use and frequency of consumption of five or more drinks consistent with one another, then these answers were taken as final.

6.2.4.5 Editing of the Frequency-of-Use Variables in Response to Data Patterns Not Involving Recency, Parent/Child Data, or Binge Alcohol Use

[Table B.3](#) in Appendix B lists detailed edits for the 12-month and 30-day frequency variables for substance use sections that contained both 12-month and 30-day frequency variables.¹⁰⁰ The edits that are described pertain to data patterns that do not involve inconsistencies (1) between the frequency data and most recent use, (2) between frequency-of-use data for parent and child data for any cocaine and crack cocaine, or (3) for the frequency of binge alcohol use in the past 30 days. The edits for frequency-of-use variables in [Table B.3](#) also pertain to any inconsistent reports that did not trigger a consistency check during the interview or situations in which a consistency check was triggered but respondents did not resolve the inconsistency.

For example, if the value for the 12-month frequency fell within the range of a respondent's estimate for the 30-day frequency (e.g., if respondents reported using on 8 days in the past 12 months and on 6 to 9 days in the past 30 days), maximum and minimum values were created for the estimated 30-day frequency. In this example where respondents reported use on 8 days in the past 12 months but estimated using the substance on "6 to 9" days in the past 30 days, use on 6 to 8 days (as opposed to 6 to 9 days) in the past 30 days would be consistent with the respondents' answer to the 12-month frequency. Information on the maximum and minimum

¹⁰⁰ These sections were alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine.

possible number of days that respondents could have used a substance in the past 30 days was used subsequently during statistical imputation to assign a final value to the 30-day frequency.

Edit rules were implemented for the inhalants and methamphetamine modules to identify and resolve additional unallowable data patterns between the 12-month and 30-day frequencies. These edits are described in [Table B.3](#). Similar to the procedures for the prescription drug edit rules (Section 6.3.13), these edits for inhalants and methamphetamine use shared code between the editing and imputation processes.

In addition, [Table B.9](#) lists edits for the 30-day frequency variables for tobacco and nicotine vaping. There were no 12-month frequency data for tobacco or nicotine vaping that could be inconsistent with the corresponding 30-day frequency data. There also were no consistency checks for the 30-day frequency variables for tobacco or nicotine vaping.

6.2.4.6 Editing of the Frequency-of-Use Variables in Response to Parent/Child Data Patterns

[Table B.4](#) lists detailed edits for the 12-month and 30-day frequency variables for the parent/child pair of cocaine and crack cocaine. For example, respondents could report use of crack cocaine on a number of days in the past 12 months that was greater than the number of days they reported using any cocaine in the past 12 months. In these situations, the higher value from the 12-month frequency for crack was assigned to the 12-month frequency for any cocaine. The source variables for the 12-month frequency for any cocaine also were edited to indicate the movement of data from the 12-month frequency of crack cocaine use to the 12-month frequency for any cocaine.

Because the hallucinogens section did not include questions on the frequency of use in the past 12 months or past 30 days for the child hallucinogens, inconsistencies in parent/child data for the frequency of use did not occur for hallucinogens.

The hallucinogens section also included follow-up questions for the 12-month frequency variables. These follow-up questions were asked if respondents originally reported that they last used any hallucinogen more than 12 months ago but subsequently reported more recent use of any hallucinogen or any of the child drugs in response to consistency checks between the recency questions for any hallucinogen and the child drugs. Respondents' original answer of use of the parent more than 12 months ago would cause them to be skipped out of the 12-month frequency-of-use questions. Therefore, when respondents gave some updated indication of use in the past 12 months, they were asked to fill in previously missing information about their frequency of use in the past 12 months.

The hallucinogens section had similar follow-up questions for the 30-day frequency of use. These questions were asked when respondents originally indicated that they last used any hallucinogen more than 30 days ago but subsequently reported that they last used any hallucinogen or any of the child drugs within the past 30 days. Thus, data on the frequency of hallucinogen use in the past 30 days were intended to be supplied by respondents (instead of

through statistical imputation) if they subsequently indicated some hallucinogen use in the past 30 days but did not originally report using hallucinogens in the past 30 days.

6.2.4.7 Editing of the Frequency-of-Use and Related Variables Involving Binge Alcohol Use

[Table B.5](#) includes edits related to the data for the frequency of binge alcohol use in the past 30 days. For example, question AL07 in the alcohol section (edited variable ALCUS30D) asked respondents to report the usual number of drinks that they consumed on the days they drank alcohol (or the single day they drank) in the past 30 days. Although this is not a frequency variable per se, information from this variable was used to edit the 30-day frequency data for alcohol and the frequency data for binge alcohol use. For example, if male respondents reported having five or more drinks per occasion on exactly the same number of days that they reported drinking any alcohol in the past 30 days, then it would logically follow that their usual number of drinks per day had to have been five or more. A similar principle applied to females who reported having four or more drinks per occasion on exactly the same number of days that they drank any alcohol in the past 30 days. If respondents who binge drank on the same number of days that they drank any alcohol in the past 30 days reported usually having fewer than four drinks (for females) or fewer than five drinks (for males) on the days when they drank alcohol in the past 30 days, ALCUS30D was assigned a special code of 975 (AT LEAST 4 OR 5 Logically assigned) to indicate usual consumption of a number of drinks that was at or above the threshold for binge alcohol use for males or females.

Similarly, if respondents drank on only 1 day in the past 30 days and reported having fewer than four drinks (for females) or five drinks (for males) on that 1 day in question AL07, but they answered question AL08 as “don’t know” or “refused,” it would logically follow that these respondents would not have been binge alcohol users in the past 30 days. When this pattern occurred, the edited variable ALCBNG30D (corresponding to question AL08) was assigned a special code of 80 (NO OCCASIONS OF 4+/5+ DRINKS IN THE PAST 30 DAYS Logically assigned) to denote that respondents logically were inferred not to have been binge drinkers in the past 30 days.

In addition, respondents who drank alcohol on more than 1 day in the past 30 days could report usual consumption of four or more drinks (for females) or five or more drinks (for males) but answer AL08 as “don’t know” or “refused.” These respondents logically would be binge drinkers on at least 1 day in the past 30 days, but the exact number of days on which they engaged in binge drinking would be unknown. No editing was done to ALCBNG30D to replace the responses of “don’t know” or “refused” with a nonmissing value. However, a flag was set to restrict donors for the imputed binge alcohol frequency to respondents who were binge alcohol users on at least 1 day in the past 30 days.

6.2.5 Initiation of Use (Age at First Use, Month of First Use, and Year of First Use)

In sections for cigarettes, smokeless tobacco, cigars, nicotine vaping, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, and methamphetamine, respondents who

reported lifetime use were asked how old they were when they first used the substance of interest. If respondents reported first using the substance within 1 year of their current age, they were asked to report the specific month and year when they first used, with response options for 2 years before the current year, 1 year before the current year, and the current year (i.e., 2021, 2022, and 2023 in the 2023 questionnaire). If respondents reported first using the substance at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the first use of the substance occurred in the current year. These respondents were asked only for the month that they first used in the current year. The remaining respondents who first used a substance within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They were then routed to a question to report on the specific month that they first used the substance in the year they had reported previously.

Because the routing logic to the different versions of the MFU and YFU questions for a given substance was mutually exclusive, composite sets of MFU and YFU variables were created from the individual unedited variables. If respondents indicated a specific year that they first used a substance, the final YFU variables were recoded to replace unedited codes with values for the specific years (e.g., 2021 through 2023). If respondents confirmed that they first used a substance at their current age and were interviewed subsequently to their birthday in the current year, a code matching the current year was assigned to the YFU. This assignment was done even if respondents did not know what month they first used in the current year or if they refused to report what month they first used in the current year. If the MFU and YFU questions had been skipped because respondents first used the substance at an age more than 1 year younger than their current age, legitimate skip codes were assigned to the final MFU and YFU variables.

Consistency checks were included if the values for the MFU and YFU were inconsistent with the AFU. Specifically, for recent initiates of a given substance, the CAI program calculated a second AFU based on the MFU and YFU by comparing these data with the respondent's date of birth. This comparison was not done if the respondents reported first use of the substance in the same month that they were born. A unique AFU could not be determined from the MFU and YFU in these situations because it was not known whether the substance use occurred before or after the respondent's birthday. Similarly, a consistency check was not triggered if respondents had missing data in either of the month or year questions, such as if respondents knew the year when they first used a substance but did not know the MFU.

In remaining situations in which respondents provided complete data for the MFU and YFU, a consistency check was triggered if the MFU and YFU suggested that respondents initiated use of the substance at an earlier or a later age than what they had previously reported. For example, a consistency check was triggered if 16-year-old respondents reported first using a substance at age 16 but then reported first using the substance in a month and year that would have meant they were 15 years old when they first used it. No editing needed to be done if respondents indicated twice in a row that the AFU that was calculated from the MFU and YFU was correct.

The CAI program updated the value for the AFU (e.g., AGE1STCG for cigarettes) to agree with the values for the MFU and YFU.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their AFU (e.g., question CG04 for cigarettes) was correct, they had an opportunity to revise the values for their YFU and their MFU. If a consistency check was triggered between the AFU and data in the MFU and YFU, the MFU and YFU were updated with any year and month data that the respondent entered in the consistency checks (e.g., CGCC21 and CGCC21a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the MFU and YFU data were picked up from the original source variables (e.g., CG04A through CG04D for any cigarette use) for use in editing, such as for the “flag and impute” edits that were described in Section 6.2.2.4 and [Table 6.2](#).

[Table B.6](#) lists edits pertaining to the initiation variables when consistency checks had been triggered because of inconsistent data between the AFU and the YFU and MFU. The default when a respondent did not resolve an inconsistency between the AFU and the MFU and YFU was to favor the AFU in subsequent editing decisions. [Table B.7](#) presents information on additional edits involving the initiation variables other than for parent/child relationships, such as situations where the AFU was inconsistent with the respondent’s current age. [Table B.8](#) presents information on edits involving parent/child relationships in the substance use initiation data. In addition to parent/child relationships that have been described previously for cocaine and hallucinogens, parent/child relationships for initiation included first use of any cigarette and initiation of daily cigarette use ([Table 6.1](#)).

6.2.6 Editing of Lifetime Daily Cigarette Use Status

Lifetime users of cigarettes were asked in question CG15 whether they ever had a period when they smoked cigarettes every day for at least 30 days. Respondents were not asked question CG15 if they had already reported that they smoked cigarettes on all 30 days in the past 30 days.

Edits associated with determining whether respondents had ever smoked cigarettes daily for at least 30 days (edited variable CIGDLYMO) are presented in [Table B.9](#). In particular, if question CG15 had been skipped because respondents smoked cigarettes on all 30 days in the past 30 days, then CIGDLYMO was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). A code of 5 was assigned rather than a legitimate skip code (Section 2.4.2) to indicate that it could be logically inferred from the skip logic that the respondent had a lifetime period of having smoked cigarettes daily for at least 30 days.

6.2.7 Edits for Tobacco Variables That Do Not Undergo Imputation

The following tobacco variables were edited but did not undergo further imputation:

- lifetime smoking of 100 or more cigarettes;
- among respondents who smoked cigarettes in the past 30 days:
 - the average number of cigarettes that they smoked per day;
 - whether the brand of cigarettes that was smoked most often was a menthol cigarette;
 - whether respondents smoked a “roll-your-own” cigarette in the past 30 days.

6.2.7.1 Editing of Lifetime Smoking of 100 or More Cigarettes

Question CG16A (edited variable CIG100LF) asked cigarette smokers whether they had ever smoked 100 or more cigarettes in their lifetime. Minimal editing was done to CIG100LF. As indicated in [Table B.9](#), respondents were not asked this question if their answers to previous questions indicated that they had smoked 100 or more cigarettes. For example, respondents who smoked cigarettes in the past 30 days were asked in question CG07 to report the number of days they smoked cigarettes in the past 30 days. Respondents who smoked cigarettes on more than 1 day in the past 30 days were asked in question CG08 to report the usual number of cigarettes they smoked on those days when they smoked cigarettes. Question CG08 was a categorical variable that gave ranges of numbers of cigarettes that were smoked per day and, where relevant, the equivalent number of packs of cigarettes (e.g., 16 to 25 cigarettes per day, or about 1 pack).

If the product of the number of days that respondents smoked cigarettes and the lower bound of the range was 100 or greater, then question CG16A was skipped. For example, if respondents smoked cigarettes on all 30 days in the past 30 days and usually smoked about a pack a day (i.e., 16 to 25 cigarettes per day), then a conservative estimate of the number of cigarettes that they smoked in the past 30 days (i.e., based on the lower bound) would be $30 \times 16 = 480$.

In this situation, respondents would not be asked question CG16A because the number of cigarettes that they logically smoked in the past 30 days was greater than 100. As indicated in [Table B.9](#), if question CG16A had been skipped because the quantity and frequency of cigarette use in the past 30 days indicated that respondents were lifetime smokers of 100 or more cigarettes, then CIG100LF was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

6.2.7.2 Editing of Miscellaneous Cigarette Use Variables for the Past 30 Days

As noted previously, respondents who reported that they smoked cigarettes in the 30 days prior to the interview were asked a series of additional questions about their cigarette use in that period. These questions covered the following topics:

- the average number of cigarettes respondents smoked per day (Section 6.2.7.2.1); and
- whether the brand of cigarettes respondents smoked most often was a menthol cigarette and whether they smoked a “roll-your-own” cigarette in the past 30 days (Section 6.2.7.2.2).

Consistent with editing procedures described previously in this chapter, these variables were assigned codes of 91 if respondents never smoked cigarettes in their lifetime. These variables were assigned codes of 93 if respondents smoked cigarettes in their lifetime but not in the past 30 days.

6.2.7.2.1 *Average Number of Cigarettes Smoked Per Day*

Respondents who smoked cigarettes in the past 30 days were asked one of two possible questions regarding the number of cigarettes they smoked per day. Respondents who smoked on only 1 day in the past 30 days were asked to report the number of cigarettes they smoked on that 1 day. Respondents who smoked on more than 1 day (or who estimated the number of days they smoked in the past 30 days) were asked to report the average number of cigarettes they smoked per day. A single, composite variable (CIG30AV) was created from these two questions using data from whatever question the respondents were asked. No further editing was done to the data from these two questions if respondents smoked cigarettes in the past 30 days.

6.2.7.2.2 *Menthol Cigarettes and “Roll-Your-Own” Cigarettes*

Past month cigarette users were routed to questions regarding whether during the past 30 days the cigarette they usually smoked was menthol (CGMENTH1) and whether they smoked part or all of a roll-your-own tobacco cigarette during this same time period (CG14). Minimal editing was done to corresponding variables CIGMENTH30 and CIGROLL30. As noted above, CIGMENTH30 and CIGROLL30 were assigned codes of 91 if respondents never smoked cigarettes in their lifetime and codes of 93 if respondents smoked cigarettes in their lifetime but not in the past 30 days.

6.2.8 Edits for Additional Marijuana Variables

In 2022, several questions were added to the marijuana section of the questionnaire or moved from other sections to the marijuana section. Questions regarding use of CBD or hemp products and questions regarding the ways in which respondents used marijuana were new for 2022. Questions regarding medical marijuana use (i.e., marijuana use recommended by a doctor or other healthcare professional) were moved from the blunts section to the marijuana section.

6.2.8.1 Editing of Lifetime and Most Recent Use of CBD or Hemp Products

Respondents were asked about lifetime and most recent use of CBD or hemp products (questions CBDLIF and CBDLAST, respectively). Minimal editing was done to create the variables CBDHMPEVR and CBDHMPREC. Respondents who reported never using CBD or hemp products (CBDHMPEVR = 2) were assigned a value of 91 (NEVER USED CBD OR HEMP PRODUCT) to the edited variable CBDHMPREC. If respondents reported lifetime use of CBD or

hemp products (CBDHMPEVR = 1), but they answered CBDLAST as “don’t know” or “refused,” then CBDHMPREC was assigned a code of 9, where 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

6.2.8.2 Editing of Medical Marijuana Use Variables

In 2022, questions MJMM01 and MJMM02 regarding medical marijuana use were moved from the blunts section to the marijuana section. Because of this change, the routing logic for the medical marijuana questions was based only on reported use of marijuana in the marijuana section, and not whether respondents smoked cigars with marijuana in them (i.e., blunts). Thus, the edited variables corresponding to questions MJMM01 and MJMM02 were renamed in 2022.

If respondents reported use of marijuana in the past 12 months in the marijuana section, they were asked question MJMM01 regarding whether any of their marijuana use in the past 12 months was recommended by a doctor or other healthcare professional. If respondents reported any medical marijuana use in the past 12 months (MJMM01 = 1), they were asked question MJMM02 regarding whether all of their marijuana use in the past 12 months was recommended by a doctor or other healthcare professional.

An important aspect of the processing of variables pertaining to medical marijuana use consisted of assigning codes of 91, 93, and 99 (Section 2.4.2) to variables that had been skipped because the medical marijuana questions did not apply to respondents. For example, if respondents never used marijuana, they were assigned a code of 91 (NEVER USED MARIJUANA) to the variables MJANYMEDYR (corresponding to MJMM01) and MJALLMEDYR (corresponding to MJMM02). Similarly, if MJMM01 and MJMM02 had been skipped because respondents last used marijuana more than 12 months ago, then MJANYMEDYR and MJALLMEDYR were assigned a code of 93 (DID NOT USE MARIJUANA IN THE PAST 12 MONTHS). If respondents were past year users of marijuana, but they answered question MJMM01 as “no,” then MJALLMEDYR was assigned a code of 99 (LEGITIMATE SKIP).

If respondents had not been asked MJMM01 and MJMM02, the respective edited variables MJANYMEDYR and MJALLMEDYR retained missing values (i.e., blank; see Section 2.2.1) in the following situations:

- The edited variable from the marijuana section for most recent use of marijuana (MJREC) had a missing value; or
- MJREC indicated use of marijuana at some point in the respondent’s lifetime but the exact period of most recent use was unknown (MJREC = 9).

Respondents who previously reported marijuana use in the past 12 months could answer question MJMM01 as “yes” (MJANYMEDYR=1), even if they were living in a state where a law approving marijuana for medical use had not passed at the time they were interviewed.

[Table 6.4](#) lists the 38 states and the District of Columbia where a bill or initiative approving marijuana for medical use was signed into law or approved by voters by the end of 2023, along with the dates when the laws were enacted in these states.

Table 6.4 States Where a Law or Initiative Allowing Marijuana for Medical Use was Passed as of December 31, 2023

State	Date Enacted (Approved, Passed, or Signed into Law)	State	Date Enacted (Approved, Passed, or Signed into Law)
California	November 5, 1996	Illinois	August 1, 2013
Alaska	November 3, 1998	Maryland	April 14, 2014
Oregon	November 3, 1998	Minnesota	May 29, 2014
Washington	November 3, 1998	New York	July 5, 2014
Maine	November 2, 1999	Pennsylvania	April 17, 2016
Hawaii	June 14, 2000	Louisiana	May 19, 2016
Colorado	November 7, 2000	Ohio	June 8, 2016
Nevada	November 7, 2000	Arkansas	November 8, 2016
Vermont	May 26, 2004	Florida	November 8, 2016
Montana	November 2, 2004	North Dakota	November 8, 2016
Rhode Island	January 3, 2006	West Virginia	April 19, 2017
New Mexico	March 13, 2007	Oklahoma	June 26, 2018
Michigan	November 4, 2008	Missouri	November 6, 2018
New Jersey	January 18, 2010	Utah	November 6, 2018
District of Columbia	May 21, 2010	South Dakota	November 3, 2020
Arizona	November 2, 2010	Virginia	March 19, 2021
Delaware	May 13, 2011	Alabama	May 17, 2021
Connecticut	May 31, 2012	Mississippi	February 2, 2022
Massachusetts	November 6, 2012	Kentucky	March 31, 2023
New Hampshire	July 23, 2013		

Two recoded variables (MEDMJPAS and MEDMJCHK2) were created that compare reported medical marijuana use with medical marijuana laws in the respondent's state of residence. These variables examined (1) reported medical marijuana use, (2) state of residence at the time of interview, (3) dates when medical marijuana laws were signed into law or approved by voters (where applicable), and (4) the interview dates. MEDMJPAS and MEDMJCHK2 were based on the dates of passage of medical marijuana laws rather than the presumed *effective* dates of laws because the effective dates could be subject to change and were subject to differing interpretations of what constituted the "effective" dates.¹⁰¹

¹⁰¹ A recoded version of MEDMJPAS (MEDMJPA2) is included on the 2023 NSDUH public use file but is not in the restricted-use data files available for analysis in NSDUH's online Data Analysis System (DAS). Documentation of the creation of variables in earlier years that made assumptions about the effective dates of medical marijuana laws can be found in the *2016 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 10: Editing and Imputation Report* (Center for Behavioral Health Statistics and Quality, 2018).

The variable MEDMJPAS was created based on the information in [Table 6.4](#) on the status of state medical marijuana laws in 2023, along with information on the date of interview and the state where respondents were living on the date they were interviewed. Thus, creation of MEDMJPAS was independent of how respondents answered questions about past year medical marijuana use.

MEDMJPAS was coded as follows:

- MEDMJPAS = 1: The respondent was in a state where a law or initiative allowing the use of marijuana for medical reasons had been passed on or before the interview date.
- MEDMJPAS = 2: The respondent was in a state in which a law or initiative allowing the use of marijuana for medical reasons had not been passed at any time during the survey year.
- MEDMJPAS = 3: The respondent was interviewed in a state where a law or initiative allowing the use of marijuana for medical reasons had been passed during the survey year but after the interview date.
- MEDMJPAS = 4: The respondent was in a state where a law or initiative allowing the use of marijuana for medical reasons was enacted at some point during the survey year but had been overturned before the interview date.

An additional related variable, MEDMJCHK2, was created to identify reports of medical marijuana use in the past 12 months from question MJMM01 (MJANYMEDYR = 1) when respondents were not living in a state in which a law or initiative allowing the use of marijuana for medical reasons had been passed at the time they were interviewed ([Table 6.4](#)).¹⁰² MEDMJCHK2 compares reports of medical marijuana use in the past 12 months with the states where respondents were living at the time they were interviewed.

Values of 1 through 6 in MEDMJCHK2 have the following meanings:

- MEDMJCHK2 = 1: The respondent reported marijuana use recommended by a doctor in the past 12 months and was living in a state in which a law or initiative allowing medical marijuana use had been passed at the time the respondent was interviewed.
- MEDMJCHK2 = 2: The respondent reported marijuana use recommended by a doctor in the past 12 months and was living in a state in which a law or initiative allowing medical marijuana use had not been passed at any time in the current survey year.
- MEDMJCHK2 = 3: The respondent reported marijuana use recommended by a doctor in the past 12 months and was living in a state in which a law or initiative allowing medical marijuana use had been passed during the current survey year but after the respondent was interviewed.
- MEDMJCHK2 = 4: The respondent used marijuana in the past 12 months but reported that use of marijuana had not been recommended by a doctor or other healthcare professional (MJANYMEDYR = 2).

¹⁰² MEDMJCHK2 is not available on the 2023 NSDUH public use file or for online analysis in the NSDUH DAS.

- MEDMJCHK2 = 5: Whether any of the respondent's past year use of marijuana was recommended by a doctor or other health professional was unknown.
- MEDMJCHK2 = 6: The respondent reported marijuana use recommended by a doctor or other healthcare professional and was living in a state where a medical marijuana law or initiative had been passed before the beginning of the survey year but had been overturned before the interview date (MEDMJPAS=4).

In addition, codes of 91 and 93 were used in MEDMJCHK2 to identify respondents who reported never using marijuana or blunts and not using marijuana or blunts in the past year, respectively.

6.2.8.3 Editing of Variables for Ways to Use Marijuana

In 2022, questions were added to the marijuana section regarding the ways in which respondents used marijuana or any cannabis product in the past 12 months and in the past 30 days. The ways asked about included (1) smoking; (2) vaping; (3) dabbing waxes, shatter, or concentrates; (4) eating or drinking; (5) putting drops, strips, lozenges, or sprays in the mouth or under the tongue; (6) applying lotion, cream, or patches to the skin; (7) taking pills; and (8) in some other way.

The questions regarding the ways in which respondents used marijuana were "enter all that apply" questions that allowed respondents to report more than one way in which they used marijuana in a given period. As noted in Section 2.4.4, these variables were coded as 1 (Yes) if a particular way was chosen from a list or as 2 (No [not entered]) if a particular way was not chosen but at least one of the other ways was chosen from the list. If respondents did not know or refused to report which ways they used marijuana, then all edited variables in the series were assigned level 94 or 97, respectively.

6.2.8.3.1 General Skip Logic

Respondents were asked question MJMODEPY regarding the ways in which they used marijuana in the past 12 months if they reported last using marijuana within the past 12 months, including use within the past 30 days or more than 30 days ago but within the past 12 months. Similarly, respondents were asked question MJMODEPM regarding the ways in which they used marijuana in the past 30 days if they reported last using marijuana within the past 30 days. Respondents who used marijuana within the past 30 days were asked both MJMODEPM and MJMODEPY.

Respondents who were not asked MJMODEPY or MJMODEPM because they never used marijuana were assigned a code of 91 (NEVER USED MARIJUANA) to the edited variables indicating the ways in which people used marijuana in the past 12 months and in the past 30 days. Respondents who were not asked MJMODEPY because they had used marijuana in their lifetime but not in the past 12 months were assigned a code of 93 (DID NOT USE MARIJUANA IN THE PAST 12 MONTHS) to the edited variables indicating the ways in which people used marijuana in the past 12 months. Similarly, respondents who were not asked MJMODEPM because they had used marijuana in their lifetime but not in the past 30 days were assigned a code of 93 (DID NOT USE MARIJUANA IN THE PAST 30 DAYS) to the edited variables indicating

the ways in which people used marijuana in the past 30 days. If respondents were not asked MJMODEPY or MJMODEPM because it was either unknown whether they ever used marijuana or unknown when they last used marijuana, then the edited variables retained a code of 98 (BLANK).

6.2.8.3.2 Logically Assigned Values Based on "OTHER, Specify" Responses

Respondents who indicated that they used marijuana or any cannabis product in some other way in the past 12 months were asked in "OTHER, Specify" question MJMDPYOa to describe the other ways they used marijuana in the past 12 months. Similarly, respondents who indicated that they used marijuana or any cannabis product in some other way in the past 30 days were asked in "OTHER, Specify" question MJMDPMOa to describe the other ways they used marijuana in the past 30 days. Respondents could provide up to five responses to each "OTHER, Specify" question they were asked.

For the set of edited 12-month variables (e.g., MJYRSMOKE), a value of 3 [Yes LOGICALLY ASSIGNED (from MJYROTHER)] was assigned if respondents did not indicate that they used marijuana in a particular way (e.g., smoking) in the past 12 months in question MJMODEPY, but they specified that they used marijuana in that way in the corresponding "OTHER, Specify" question MJMDPYOa. For example, if respondents did not select "Eating or drinking" from the list of response options for the past year, but they described using edibles in MJMDPYOa, then MJYREAT was assigned a code of 3.

Similarly, for the set of 30-day edited variables (e.g., MJMONSMOKE), a code of 3 [Yes LOGICALLY ASSIGNED (from MJMONOTHER)] was assigned if respondents did not indicate that they used marijuana in a particular way in the past 30 days in question MJMODEPM, but they specified using marijuana in that way in the corresponding "OTHER, Specify" question MJMDPMOa. For example, if respondents did not select "Eating or drinking" from the list of response options for the past month, but they described using edibles in MJMDPMOa, then MJMONEAT was assigned a code of 3.

6.2.8.3.3 Logically Assigned Values Because of Additional Inconsistent Responses

As noted previously, respondents who reported last using marijuana during the past 30 days were asked the questions about ways they used marijuana in both the past 12 months and the past 30 days. The following patterns of inconsistent data could occur when respondents reported marijuana use in the past 30 days:

- Respondents reported some ways of use in the past 30 days that they did not report for the past 12 months. Logically, if respondents used marijuana in a particular way in the past 30 days, then they also used marijuana in that same way in the past 12 months.
- Respondents who reported using marijuana in a particular way in the past 30 days answered the 12-month question as "don't know" or "refused."

In these situations, the edited 12-month variable corresponding to the way in which people used marijuana was assigned a level of 5 [(Yes LOGICALLY ASSIGNED (from MJMONxxxx)].¹⁰³ For example, if respondents reported smoking marijuana as the only way they used marijuana in the past 30 days, but vaping was the only way they reported using marijuana in the past 12 months, then the edited variable for smoking marijuana in the past 12 months (MJYRSMOKE) was assigned a code of 5. Therefore, these respondents would be classified as smoking marijuana in the past 30 days and smoking and vaping marijuana in the past 12 months. In this example, however, respondents were not logically inferred to have vaped marijuana in the past 30 days because they could have last vaped marijuana more than 30 days ago but within the past 12 months.

In addition, respondents who last used marijuana in any way in the past 30 days could have reported they used marijuana in a particular way in the past 12 months, but they answered the 30-day question as “don’t know” or “refused.” The following edits were made in this situation:

- If respondents reported only one way of use in the past 12 months in MJMODEPY, then it was inferred that this was the only way they could have used marijuana in the past 30 days.
 - The variable for the past 30 days that corresponded to the way in which respondents reported using marijuana in the past 12 months was assigned a code of 5 (Yes LOGICALLY ASSIGNED).
 - Variables for the past 30 days that corresponded to ways respondents did not report using marijuana in the past 12 months were assigned a code of 4 [No (Not entered) LOGICALLY ASSIGNED].
- If respondents reported more than one way of use in the past 12 months in MJMODEPY, then it could not be determined whether respondents also used marijuana in these specific ways in the past 30 days. However, it could be inferred that respondents did not use marijuana in the past 30 days in ways that they did not report for the past 12 months.
 - Variables for the past 30 days that corresponded to ways respondents did not report using marijuana in the past 12 months were assigned a code of 4 [No (Not entered) LOGICALLY ASSIGNED].
 - Responses of “don’t know” or “refused” were retained in variables for ways of use in the past 30 days that corresponded to ways in which respondents used marijuana in the past 12 months.

¹⁰³ The description of level 5 varies depending on the edited variable. For example, the description of level 5 for edited variable MJYRSMOKE (used marijuana by smoking in the past 12 months) is “[Yes LOGICALLY ASSIGNED (from MJMONSMOKE)].

6.3 Editing Procedures for the Prescription Drug Variables

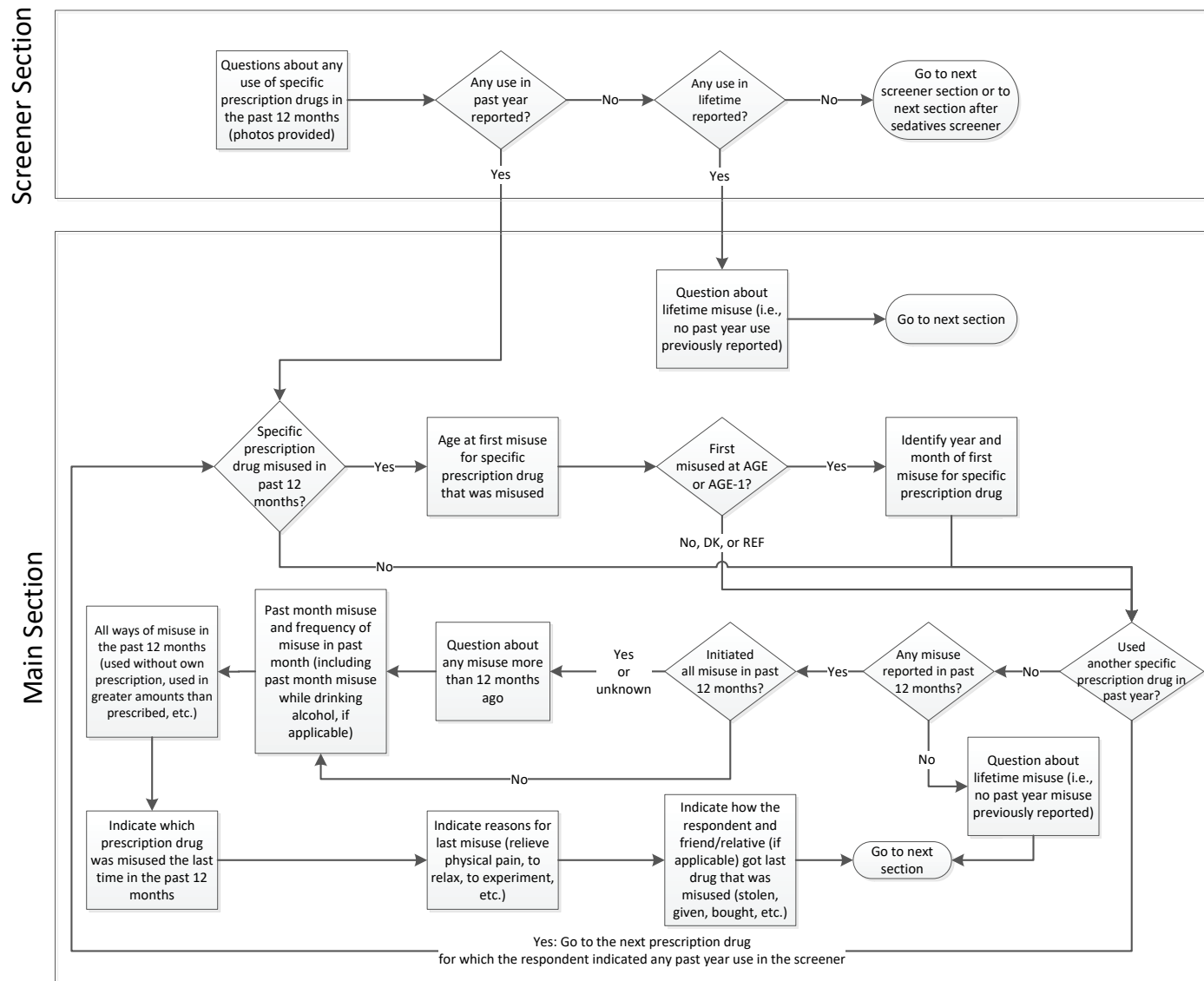
[Figure 6.2](#) provides an overview of the general routing logic for the prescription psychotherapeutic drug sections. For each category of psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), respondents first were asked “screener” questions about any use of specific prescription drugs in the past 12 months or, if they did not report use in the past 12 months, lifetime use of any prescription drug in that category.

Respondents who reported any use of prescription drugs in either the past 12 months or their lifetime were asked more detailed questions about their misuse of prescription drugs in that category (see Section 6.3.4 for the definition of misuse). The questions for misuse included the following:

- misuse of specific prescription drugs in the past 12 months, if respondents reported any use of a given specific prescription drug in that period;
- initiation of misuse of specific prescription drugs that were misused in the past 12 months (i.e., age at first misuse [AFM], and for recent initiates, the month of first misuse [MFM] and year of first misuse [YFM]);
- lifetime misuse of any prescription drug in a given category (if misuse in the past 12 months was not reported, or if respondents reported any use of prescription drugs in that category in their lifetime but not in the past 12 months);
- misuse of any prescription drugs in a given category more than 12 months prior to the interview, if respondents reported initiation of misuse in the past 12 months for all prescription drugs that they misused in that period (i.e., to identify respondents who initiated misuse of any prescription drug in the category more than 12 months prior to the interview); and
- frequency of misuse in the past 30 days.

In addition, respondents who reported misuse of any prescription psychotherapeutic drug in a given category in the past 30 days and who also drank alcohol in the past 30 days were asked whether they misused prescription drugs at the same time or within a couple of hours of when they drank alcohol in that period. Respondents who reported misuse of any prescription psychotherapeutic drug in a given category in the past 12 months were asked to report (1) the ways in which they misused prescription drugs within that category in the past 12 months (e.g., without a prescription of their own, in greater amounts than directed), (2) the last prescription drug in that category that they misused in the past 12 months (if they reported misuse of more than one prescription drug), (3) their reasons for misusing that last prescription drug they misused in the past 12 months, and (4) how they (or their friends or relatives, if applicable) obtained the last prescription drug they misused in that period. In addition, respondents who reported the misuse of prescription stimulants in the past 12 months were asked whether they used a needle to inject stimulants in the past 12 months or (if applicable) the past 30 days.

Figure 6.2 Routing Logic for Respondents in the Prescription Psychotherapeutic Drug Sections



DK = don't know; REF = refused.

6.3.1 Mentions of Over-the-Counter Drugs

To make the questions about the use of specific prescription drugs exhaustive for a given prescription psychotherapeutic drug category, respondents were asked whether they used “any other” prescription drug in that category in the past 12 months. Respondents who reported use in the past 12 months of other prescription drugs in a given category were asked whether they misused these other drugs in that period, and if so, they were asked to specify the names of up to five prescription drugs in that category that they misused in the past 12 months (Section 2.3.1.1).

One way that people can misuse prescription drugs is by taking them without having their own prescription. Because over-the-counter (OTC) drugs by definition are available without a prescription, respondents were instructed not to include OTC drugs when answering the questions in screener sections for pain relievers, stimulants, and sedatives. These drug categories include products that are available in the United States without a prescription (e.g., the pain reliever aspirin, the stimulant Dexatrim®, and the sedative Nytol®). However, in the “OTHER, Specify” data for other prescription drugs that respondents misused in the past 12 months, respondents could report the misuse of OTC drugs.

Instructions were provided for the misuse of “any other” prescription pain reliever, stimulant, or sedative to remind respondents not to include OTC medication (e.g., Tylenol®, 5-Hour Energy®, Sominex®). These instructions were not shown for the misuse of any other prescription tranquilizer because no tranquilizers are available in the United States without a prescription.

A code of 4 (No LOGICALLY ASSIGNED) was assigned to the lead question for the misuse of any other drug in that category in the past 12 months (e.g., PNR0TNMYR, corresponding to question PRYOTH for pain relievers) when the only drugs¹⁰⁴ that respondents specified for misuse in the past 12 months were OTC drugs, and (1) they reported the past year misuse of at least one other prescription drug in that section or (2) they answered questions for the misuse of other specific prescription drugs in that category in the past 12 months as “don’t know” or “refused.” That is, it could be inferred that the respondents had not misused other prescription drugs in that category in the past 12 months, even if they had misused some specific prescription drugs in that period (or had potentially misused prescription drugs for which they answered the questions for misuse in the past 12 months as “don’t know” or “refused”). Codes in the “OTHER, Specify” variables were overwritten with a code of 9989 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2) to indicate that respondents logically should have skipped these questions.

If it could be determined unambiguously that “other” prescription drugs in a given psychotherapeutic category (e.g., pain relievers) were the only prescription drugs that respondents misused in the past 12 months and the misuse of only OTC drugs was reported,

¹⁰⁴ Respondents were considered to have specified the misuse of only OTC drugs if the “OTHER, Specify” data contained reports of only OTC drugs or the data included only OTC drugs and codes for “bad data.” This condition was not considered to have been satisfied in situations where the “OTHER, Specify” data included responses of “don’t know” or “refused” in addition to reports of OTC drugs (i.e., but no reports of the misuse of prescription drugs). These responses of “don’t know” or “refused” were interpreted to mean that the respondents had potentially misused other drugs that require a prescription.

then respondents were inferred not to have misused any prescription drug in that category in the past 12 months. Respondents were misusers of only the “other” prescription drugs in a category in the past 12 months if they reported misuse of “any other” prescription drug in the category and

- they reported use in the past 12 months of specific prescription drugs, and they answered all of the follow-up questions for misuse of these specific drugs as “no,” or
- the only prescription drugs in a given category that they used in the past 12 months for any reason were other prescription drugs, such that the lead question (e.g., PRYOTH for pain relievers) asked whether respondents had misused any prescription drug in that category in the past 12 months.

Editing of other variables within a prescription drug section is discussed in further detail in subsequent sections for situations in which respondents were determined not to have misused any prescription drug in a given category because of their reports of misuse of only OTC drugs. However, no editing was done to the corresponding screener question about any use of other drugs in that category in the past 12 months (e.g., PRANYOTH for pain relievers), which resulted in respondents being routed to the question about misuse of any other drug in the category. Even if the only drugs in a given category for which respondents reported misuse in the past 12 months were OTC drugs, respondents could have used other prescription drugs in the past 12 months that they did not misuse.

6.3.2 Any Past Year or Lifetime Use of Specific Prescription Drugs

Questions for the use of specific prescription drugs in the past 12 months were “enter all that apply” questions, in which respondents could report the use of more than one drug in the list. Respondents were instructed to select “95” if they had never used any of the specific drugs in a given list in the past 12 months (e.g., the pain relievers Vicodin®, Lortab®, Norco®, Zohydro® ER, and generic hydrocodone in question PR01). Respondents could choose one or more drugs from a list or they could select “95,” but they could not do both. Data for the use or nonuse of specific prescription drugs were captured in separate variables (e.g., the variables VICOANYR, LTABANYR, NRCOANYR, ZOHYANYR, and HYCDANYR, corresponding to the specific prescription pain relievers in PR01).

Questions for the use of any other prescription drug in a given category in the past 12 months (e.g., question PRANYOTH for pain relievers) were “yes/no” questions, where 1 = Yes, and 2 = No. Respondents could choose only one response in these questions for the use of other prescription drugs in that category (i.e., besides the specific prescription drugs that respondents were asked about).

6.3.2.1 Editing of Screener Variables for Past Year Users

This section describes editing procedures for the prescription drug screener variables if respondents reported the past year use of at least one prescription drug in that section (e.g., pain relievers), including the use of any other drug in the category. Minimal editing was done to the variables for the use of other prescription drugs in a given category when

respondents reported the use of any prescription drug in a given screener section, other than to assign descriptive names to the edited variables (e.g., edited variable PNROTANYR, corresponding to question PRANYOTH). Consequently, the edited variables for the past year use of other prescription drugs in a category have values of 1 (Yes) or 2 (No) or codes for missing data (i.e., 94 for “don’t know,” 97 for “refused,” or 85 for “bad data”).¹⁰⁵

Consistent with the procedures described in Section 2.4.4, if respondents reported using one or more specific prescription drugs in the past 12 months from a given list of “enter all that apply” questions, the variables corresponding to the specific prescription drugs that respondents used were coded as 1 (Yes). Variables corresponding to other drugs in the list that respondents did not report using in the past 12 months were coded as 2, where 2 = No (not entered).

As described in Section 2.4.5, if respondents did not report using a specific prescription drug in the past 12 months but they specified misusing it in that period in the “OTHER, Specify” data for the relevant prescription drug category (e.g., pain relievers), then the corresponding screener variable for that drug was assigned a code of 5 (Yes LOGICALLY ASSIGNED). For example, if respondents did not report using the pain reliever Norco[®] in the past 12 months but reported misusing that drug in the same period, then they logically used Norco[®] for any reason in the past 12 months. The edited variable NRCOANYR for any use of Norco[®] in that period was assigned a code of 5. The following additional edits were implemented to the other screener variables from a given list when a code of 5 was assigned through these editing procedures:

- If respondents reported the use of other prescription drugs from that list in the past 12 months, then the remaining prescription drug variables from that list were coded as 1 or 2, as per the procedures described above. For example, if NRCOANYR was assigned a code of 5 but the respondent had previously reported using Vicodin[®] in the past 12 months, then VICOANYR was assigned a code of 1, and the remaining variables LTABANYR (for Lortab[®]), ZOHYANYR (for Zohydro[®] ER), and HYCDANYR (for generic hydrocodone) were assigned codes of 2.
- If respondents selected 95 to indicate that they had not used any of the prescription drugs in that list in the past 12 months, then the edited variables for the remaining prescription drugs in that list were assigned a code of 2 (No [Not entered]). For example, if NRCOANYR was assigned a code of 5 but the respondent had previously reported not using any of the pain relievers in question PR01, then the remaining variables VICOANYR, LTABANYR, ZOHYANYR, and HYCDANYR were assigned codes of 2.

¹⁰⁵ As noted in Section 2.3.2.1, prescription drug screener data were set to “bad data” if responses were identified as patterned.

- If respondents answered an entire list as 94 (“don’t know”) or 97 (“refused”) but they specified one of the drugs in the list as some other prescription drug that they misused in the past 12 months, then the edited variable for that prescription drug was assigned a code of 5 and the variables for the remaining prescription drugs in the list retained the codes of 94 or 97. For example, if NRCOANYR was assigned a code of 5 but the respondent had originally answered question PR01 as “refused,” then the remaining variables VICOANYR, LTABANYR, ZOHYANYR, and HYCDANYR retained codes of 97.

The additional edits described below were implemented if respondents did not report the use of any specific prescription drugs in the past 12 months from a given list in an “enter all that apply” question (e.g., PR01) but they reported use of other prescription drugs in that category in the past 12 months elsewhere in that screener section. These edits applied to situations in which no screener variables corresponding to prescription drugs in a given list were assigned a code of 5 based on “OTHER, Specify” data for the misuse of prescription drugs in that category.

- If respondents selected 95 to indicate that they had not used any prescription drug in a given list in the past 12 months, then all of the edited variables corresponding to the prescription drugs in that list were assigned a code of 95, where 95 = I have not used any of these [DRUGS (e.g., pain relievers)] in the past 12 months. For example, if respondents selected 95 for question PR01 to indicate that they had not used any of these pain relievers in the past 12 months, but they reported in question PR02 that they used OxyContin® in that period, then the edited variables VICOANYR, LTABANYR, NRCOANYR, ZOHYANYR, and HYCDANYR were still assigned a code of 95.
- If respondents reported that they did not know or they refused to report whether they used any prescription drug in a given list in the past 12 months, then all of the edited variables corresponding to the prescription drugs in that list were assigned codes of 94 (“don’t know”) or 97 (“refused”), respectively. For example, if respondents did not know whether they used any of the pain relievers in question PR01 in the past 12 months, then the edited variables VICOANYR, LTABANYR, NRCOANYR, ZOHYANYR, and HYCDANYR were assigned a code of 94.

When respondents reported the use of at least one prescription drug in a given category in the past 12 months, the question for the lifetime use of any prescription drug in that category was skipped (e.g., question PRLANY for pain relievers) because users in the past 12 months logically had used prescription drugs in that category in their lifetime. The corresponding edited variable for lifetime use (e.g., PNRANYLIF, corresponding to PRLANY) was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

6.3.2.2 Editing of Screener Variables for Lifetime but Not Past Year Users

Respondents who did not report using any prescription drug in a given category in the past 12 months (including respondents who did not know or refused to report whether they used certain prescription drugs in that period) were asked whether they ever used any prescription drug in that category for any reason (e.g., question PRLANY for pain relievers). When respondents answered the question for any use in their lifetime as “yes,” then the edited

variable (e.g., PNRANYLIF, corresponding to PRLANY) was assigned a code of 1 (Yes). The edits described below were implemented when respondents reported lifetime use of prescription drugs in a given category for any reason (e.g., PNRANYLIF = 1).

- If respondents selected 95 for all “enter all that apply” questions in the screener section to indicate that they had not used any of these prescription drugs in the past 12 months, and they also reported that they did not use any other drug in that category in the past 12 months, then all edited screener variables pertaining to the past 12 months were assigned a code of 93, where 93 = DID NOT USE [PRESCRIPTION DRUG CATEGORY] IN THE PAST 12 MONTHS. For example, if PNRANYLIF = 1, responses of 95 were selected in questions PR01 through PR10, and PRANYOTH was answered as 2 (no), then the edited variables VICOANYR through PNOTANYR (corresponding to PRANYOTH) were assigned a code of 93 (DID NOT USE PAIN RELIEVERS IN THE PAST 12 MONTHS).
- If respondents had a combination of responses of 95 or missing data (i.e., codes of 85, 94, or 97 for “bad data,” “don’t know,” or “refused,” respectively), then the edited variables corresponding to prescription drugs that respondents indicated that they explicitly did not use in the past 12 months were assigned a code of 95 (for nonuse of any of these prescription drugs in the past 12 months). The other variables with missing data retained the codes for missing values because the respondents could have used some of these prescription drugs in the past 12 months. For example, if respondents selected 95 in question PR01 but they answered question PR02 as “don’t know,” then the edited variables VICOANYR, LTABANYR, NRCOANYR, ZOHYANYR, and HYCDANYR from PR01 were assigned a code of 95. The edited variables OXCNANYR (for OxyContin®), PRCTANYR (for Percocet®), and PRDNANYR (for Percodan®) were assigned a code of 94 for “don’t know.”

6.3.2.3 Editing of Screener Variables for Lifetime Nonusers

When respondents did not report use in the past 12 months of any prescription drug within a given category, and they answered the question for any use in their lifetime as “no,” then the edited variable (e.g., PNRANYLIF, corresponding to PRLANY) was assigned a code of 2 (No). The edits described below were implemented when respondents reported that they never used any prescription drug in a given category for any reason (e.g., PNRANYLIF = 2).

- If respondents selected 95 for all “enter all that apply” questions in the screener section to indicate that they had not used any of these prescription drugs in the past 12 months, and they also reported that they did not use any other drug in that category in the past 12 months, then all edited screener variables pertaining to the past 12 months were assigned a code of 91, where 91 = NEVER USED [PRESCRIPTION DRUG CATEGORY]. For example, if PNRANYLIF = 2, responses of 95 were selected in questions PR01 through PR10, and if PRANYOTH was answered as 2 (no), then the edited variables VICOANYR through PNOTANYR (corresponding to PRANYOTH) were assigned a code of 91 (NEVER USED PAIN RELIEVERS).

- If respondents had a combination of responses of 95 or missing data (i.e., codes of 85, 94, or 97 for “bad data,” “don’t know,” or “refused,” respectively), then the edited variables corresponding to prescription drugs that respondents indicated that they explicitly did not use in the past 12 months were assigned a code of 91. If variables had missing data for use in the past 12 months, the missing values were replaced with codes of 81 (see the example below for pain relievers). Thus, the editing procedures gave precedence to reports of lifetime nonuse of any prescription drug in a category over missing data for use in the past 12 months. For example, if respondents selected 95 in question PR01 but they answered question PR02 as “don’t know,” then the edited variables VICOANYR, LTABANYR, NRCOANYR, ZOHYANYR, and HYCDANYR from PR01 were assigned a code of 91. The edited variables OXCNANYR (for OxyContin®), PRCTANYR (for Percocet®), and PRDNANYR (for Percodan®) were assigned a code of 81 (NEVER USED PAIN RELIEVERS Logically assigned).

6.3.3 Most Recent Use of Prescription Drugs for Any Reason

Unlike the sections of the interview for substances other than prescription drugs, respondents were not explicitly asked to report the last time when they used prescription drugs for any reason. Instead, recoded variables were created for the most recent use of prescription drugs in a given category for any reason. Creation of the recoded variable PNRANYREC for the most recent use of prescription pain relievers for any reason is described in the examples below.

- If use in the past 12 months was reported for at least one prescription pain reliever in the screener section, then PNRANYREC for the most recent use was set to 1, where 1 = Within the past 12 months.
- If no use of specific prescription pain relievers in the past 12 months was reported in the screener section, but lifetime use of any prescription pain reliever was reported, then values were assigned to PNRANYREC as follows:
 - PNRANYREC was set to 2 (More than 12 months ago) if all edited screener variables for the use of pain relievers in the past 12 months had been set to 93 (DID NOT USE PAIN RELIEVERS IN THE PAST 12 MONTHS). When lifetime users explicitly reported that they had not used any pain relievers in the past 12 months, the most recent use of any pain reliever logically occurred more than 12 months prior to the interview.
 - PNRANYREC was set to 9 (Used at some point in the lifetime) if any of the edited screener variables for the past 12 months had missing values (e.g., “don’t know”). Lifetime users in this situation could have used pain relievers in the past 12 months but did not report doing so.
- If question PRLANY for any lifetime use of pain relievers was answered as “no,” then PNRANYREC was set to 91 (i.e., never used).

- If PRLANY was answered as “don’t know” or “refused,” responses of 95 were selected in questions PR01 through PR10, and PRANYOTH was answered as 2 (no), then PNRANYREC was set to 83 (DID NOT USE IN THE PAST 12 MONTHS [LIFETIME UNKNOWN] Logically assigned). Respondents in this situation definitely did not use pain relievers in the past 12 months, but it was unknown whether they used them more than 12 months ago or they never used pain relievers.
- Otherwise, PNRANYREC was assigned a code of 98 (blank) to indicate that any lifetime use or lifetime nonuse was unknown. This pattern could occur in the following situations:
 - The question for any lifetime use was answered as “don’t know” or “refused,” and responses of “don’t know” or “refused” existed in the screener items for the past 12 months (i.e., but without any reports of use in the past 12 months).
 - The corresponding pain reliever screener variables for all questions that respondents answered had been assigned a code for bad data because of patterned responses (Section 2.3.2.1).

6.3.4 Misuse of Specific Prescription Drugs in the Past 12 Months

As shown in [Figure 6.2](#), respondents who reported any use of specific prescription drugs in the past 12 months were asked whether they misused these prescription drugs in that period. Misuse is defined as the use of prescription drugs in any way a doctor did not direct respondents to use them. Examples of behaviors constituting misuse that were presented to respondents included (1) use without a prescription of the respondent’s own; (2) use in greater amounts, more often, or longer than told to take a drug; or (3) use in any other way a doctor did not tell respondents to take a drug.

6.3.4.1 Editing of Specific Prescription Drug Misuse Variables for Prescription Drug Misusers in the Past 12 Months

Questions for the misuse of specific prescription drugs in the past 12 months (e.g., question PRY01 for the misuse of Vicodin®) were “yes/no” questions, where 1 = Yes, and 2 = No. In the absence of any patterned responses for a given prescription drug section (Section 2.3.2), edited variables for misuse of specific prescription drugs would be coded as 1 (“yes”), 2 (“no”), 94 (“don’t know”), or 97 (“refused”) for the prescription drugs that respondents used for any reason in the past 12 months. Questions for the misuse of prescription drugs that respondents did not report using in the past 12 months were coded as 93 (or 993, etc.) to indicate that the respondents did not use those drugs in the past 12 months. For example, if respondents reported the past year use of Vicodin® in question PR01, did not report using any of the other drugs listed in that question, and there were no reports of the misuse of these other drugs in the past 12 months in the “OTHER, Specify” data, then the edited variables LTABNMYR (for the misuse of Lortab®), NRCONMYR (for the misuse of Norco®), ZOHYNMYR (for the misuse of

Zohydro® ER), and HYCDNMYR (for the misuse of generic hydrocodone) were assigned codes of 93.¹⁰⁶

As described in Section 2.4.5, if respondents did not report that they misused a specific prescription drug in the past 12 months, but they specified misusing it in that period in the "OTHER, Specify" data for the relevant prescription drug category (e.g., pain relievers), then the corresponding variable for the misuse of that drug in the past 12 months was edited as follows:

- If the respondents reported any use of that drug in the past 12 months in the corresponding screener section but did not answer the question for misuse as "yes," then the edited misuse variable was assigned a code of 3 (Yes LOGICALLY ASSIGNED). For example, suppose respondents reported using the pain reliever Norco® in the past 12 months. For these respondents, the screener variable NRCOANYR for Norco® would be coded as 1 (Yes). If the respondents did not report misusing the drug but specified it as some other pain reliever that they misused in the past 12 months, then NRCONMYR (corresponding to question PRY03) was assigned a value of 3.
- If the screener variable (e.g., NRCOANYR for Norco®) had been assigned a code of 5 because respondents had not reported using the drug in the past 12 months, but they specified its misuse in that period in the "OTHER, Specify" data (Section 6.3.2.1), then the edited misuse variable for that drug (e.g., NRCONMYR) also was assigned a code of 5 (Yes LOGICALLY ASSIGNED).

Respondents who reported the misuse of at least one prescription drug in a given category in the past 12 months were not asked whether they ever misused any prescription drug in that category (e.g., question PRL02) because these respondents logically had misused prescription drugs in that category in their lifetime. The corresponding edited variable for lifetime misuse (e.g., PNRNMLIF for pain relievers) was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

6.3.4.2 Editing of Specific Prescription Drug Misuse Variables for Misusers of Over-the-Counter Drugs in the Past 12 Months

As noted in Section 6.3.1, respondents were logically inferred not to have misused other prescription drugs in a given category when the only drugs that they specified for misuse in the past 12 months were OTC drugs. No further editing was done to the variables for the past year misuse of specific prescription drugs if the lead question for the misuse of any other drug in that category in the past 12 months (e.g., PNROTMYR, corresponding to question PRYOTH for pain relievers) had been assigned a code of 4 (No LOGICALLY ASSIGNED) but the respondent reported misuse in the past 12 months of other prescription drugs in that category.

¹⁰⁶ Responses of "don't know" or "refused" for specific prescription drugs or "any other" drug in a category were considered to be potential reports of use in the past 12 months.

Respondents were inferred not to have misused any prescription drug in the overall prescription drug category (e.g., pain relievers) in the past 12 months in the following situations where the respondents reported the misuse of only OTC drugs:

- All other questions for the misuse of specific prescription drugs in the past 12 months had been answered as “no” or had been legitimately skipped because respondents did not report any use of that drug in the same period.
- The only screener question for which respondents reported any use in the past 12 months was for “any other” drug in the category, such that the corresponding question for misuse (e.g., PRYOTH for pain relievers) asked about the misuse of any prescription drug in that category (i.e., as opposed to the misuse of any other prescription drug in that category).

Assignment of the relevant code to the corresponding recoded variable for the most recent misuse when respondents misused only OTC drugs is described in Section 6.3.5.

No additional editing was done if respondents had been inferred not to have misused other prescription drugs in a given category in the past 12 months because the only drugs they misused in that period were OTC drugs, and the respondents answered questions for the misuse of other specific prescription drugs in that period as “don’t know” or “refused.” In this situation, respondents had reported any use of specific prescription drugs in the past 12 months in addition to reporting any use of “any other” prescription drug in that category. Consequently, these respondents could have misused the specific prescription drugs for which they answered the misuse questions as “don’t know” or “refused.”

6.3.4.3 Editing of Specific Prescription Drug Misuse Variables for Lifetime but Not Past Year Misusers

Respondents who did not report any past year use of prescription drugs in the relevant screener section but who reported lifetime use for any reason of any prescription drug in that category were asked whether they ever misused any prescription drug in that category (e.g., question PRL01 for pain relievers). Respondents who were asked about misuse of specific prescription drugs in the past 12 months but who did not report misuse in that period also were asked whether they ever misused any prescription drug in that category (e.g., question PRL02 for pain relievers). Thus, these lifetime misuse questions were mutually exclusive. For pain relievers, for example, respondents could be routed to either PRL01 or PRL02 to indicate whether they had ever misused pain relievers in their lifetime, but they could not be routed to both questions.

If respondents reported lifetime misuse (e.g., PNRNMLIF = 1 for pain relievers, from PRL01 or PRL02), then the following edits were implemented for the variables for the misuse of specific prescription drugs:

- If all questions that respondents were asked for their misuse of specific prescription drugs in the past 12 months were answered as “no,” then all variables for the misuse of specific prescription drugs in the past 12 months were assigned a code of 93 (or 993, etc.) to indicate that the respondents misused no prescription drugs within that psychotherapeutic category in the past 12 months.
- If respondents did not report the misuse of prescription drugs within a psychotherapeutic category in the past 12 months but they had missing data for the misuse of other prescription drugs within that category, then the following edits were implemented:
 - Variables corresponding to prescription drug questions that were skipped because respondents did not report any use in the past 12 months in the corresponding screener section were assigned a code of 93 (or 993, etc.).
 - If respondents were routed to a question for a specific prescription drug, and that question was answered as “no,” then the corresponding edited variable (e.g., NRCONMYR for the misuse of Norco[®] in the past 12 months) retained a value of 2 (i.e., No).
 - Variables for the misuse of specific prescription drugs in the past 12 months that had missing data retained missing values.

6.3.4.4 Editing of Specific Prescription Drug Variables for No Lifetime Misuse

Respondents who answered a question for lifetime misuse as “no” (e.g., question PRL01 or PRL02 for pain relievers) were determined never to have misused any prescription drug in that category in their lifetime. The edits described below were implemented when respondents were routed to the section for the misuse of prescription drugs in a given category (e.g., pain relievers) and they reported that they never misused any prescription drug in that category (e.g., PNRNMLIF = 2).

- If the variables for the misuse in the past 12 months of all specific prescription drugs in that section were coded as 2 (i.e., no) or 93, then all edited variables for the misuse of these specific prescription drugs were coded as 91 (or 991, etc.) to indicate no lifetime misuse.
- If respondents had a combination of responses of 2, 93, or missing data, then the edited variables corresponding to prescription drugs that respondents explicitly did not misuse were assigned a code of 91 (or 991, etc.). If variables had missing data for misuse in the past 12 months, then any answers to these questions that were coded with missing values (e.g., “don’t know”) were replaced with codes of 81 (or 981, etc.). Thus, the editing procedures gave precedence to reports of no lifetime misuse of any prescription drug in a category over missing data for use in the past 12 months.

6.3.5 Most Recent Misuse of Prescription Drugs

Similar to the variables for the most recent use of prescription drugs in a given category for any reason, recoded variables were created for the most recent misuse of any prescription drug in that category. Creation of the recoded variable PNRNMREC for the most recent misuse of prescription pain relievers is described as an example.

- If misuse of any prescription pain reliever in the past 30 days was reported in question PRM01, or if PRM01 had been skipped because respondents reported that they initiated the misuse of one or more pain relievers in the 30 days prior to the interview (Section 6.3.6), then PNRNMREC was set to 1, where 1 = In the past 30 days.
- If respondents misused one or more specific pain relievers in the past 12 months, then values were assigned to PNRNMREC as follows:
 - PNRNMREC was set to 2 (More than 30 days ago but within the past 12 months) if PRM01 was answered as “no.” In this situation, respondents who reported misuse in the past 12 months unambiguously had not misused any prescription pain reliever in the past 30 days.
 - PNRNMREC was set to 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) if PRM01 was answered as “don’t know” or “refused.” These respondents definitely misused prescription pain relievers in the past 12 months and also potentially misused pain relievers in the past 30 days.
- If no misuse of specific prescription pain relievers in the past 12 months was reported, but lifetime misuse of any prescription pain reliever was reported, then values were assigned to PNRNMREC as follows:
 - PNRNMREC was set to 3 (Misused more than 12 months ago) if all questions for the misuse of specific pain relievers in the past 12 months were answered as “no” or were coded as 93 because they had been legitimately skipped.
 - PNRNMREC was set to 9 (Misused at some point in the lifetime) if any of the questions for misuse of specific pain relievers in the past 12 months were answered as “don’t know” or “refused.” Respondents in this situation could have misused pain relievers in the past 12 months but did not report doing so. These respondents also could have misused pain relievers in the past 30 days but were not asked PRM01.
- If the only prescription drugs in a given category that respondents misused in the past 12 months were OTC drugs, then PNRNMREC was set to 83 (DID NOT MISUSE IN THE PAST 12 MONTHS (LIFETIME UNKNOWN) Logically assigned). Because the respondents during the interview reported misuse in the past 12 months, the lifetime misuse question (e.g., PRL02 for pain relievers) was skipped. Therefore, these respondents could have last misused pain relievers more than 12 months ago or never misused them.
- If questions PRL01 or PRL02 for any lifetime misuse of pain relievers were answered as “no,” then PNRNMREC was set to 91 (i.e., never used). The answer of “no” in the lifetime question overruled any answers of “don’t know” or “refused” for the misuse of specific prescription drugs in the past 12 months.

- If the screener data indicated that respondents had never used pain relievers in their lifetime for any reason (PNRANYREC = 91), then PNRNMREC also was set to 91.
- Otherwise, PNRNMREC was assigned a code of 98 (blank) to indicate that lifetime misuse or no lifetime misuse was unknown. This pattern could occur in the following situations:
 - Either question for lifetime misuse was answered as “don’t know” or “refused.” If respondents were routed to question PRL02 because they reported any use of pain relievers in the past 12 months, PNRNMREC was set to 98 when PRL02 was answered as “don’t know” or “refused,” and responses of “don’t know” or “refused” existed in the items for misuse in the past 12 months (i.e., but without any reports of misuse in the past 12 months).
 - All pain reliever screener and misuse data had been assigned a code for bad data because of patterned responses (Section 2.3.2.1).

6.3.6 Initiation of Misuse of Prescription Drugs

The questions for the initiation of misuse focused only on the specific prescription drugs that respondents misused within the past year. Respondents were not asked about initiation for prescription drugs that they may have misused in their lifetime but not in the past 12 months. Consequently, respondents cannot report initiation in the past 12 months without also reporting misuse in that period.

6.3.6.1 Identification of the Initiation of Misuse in the Past 12 Months

To establish that respondents had initiated the misuse of all prescription drugs in a given category in the past 12 months, any initiation prior to that period must be ruled out. Even if respondents reported initiation in the past 12 months for all of the specific prescription drugs in a given overall category that they misused in that period, this information does not conclusively rule out the misuse of other prescription drugs that may have occurred more than 12 months prior to the interview. For example, suppose respondents reported the misuse of one pain reliever (e.g., OxyContin®) in the past 12 months and that they first misused this pain reliever in the same period. This information established that these respondents initiated the misuse of OxyContin® in the past 12 months. However, this information is not sufficient for establishing that the respondents initiated the misuse of all pain relievers in the past 12 months. For example, some respondents may have misused other pain relievers (e.g., Vicodin®) more than 12 months ago but were not asked about the misuse of these other pain relievers. These respondents would not be past year initiates for the misuse of any prescription pain reliever.

Therefore, respondents who reported that they initiated misuse in the past 12 months for all of the specific prescription drugs in a given category that they misused in that period were asked a follow-up question (e.g., question PRL03 for pain relievers) to establish whether they had *ever* misused prescription drugs in that category more than 12 months before being interviewed. Respondents also were asked this follow-up question if there were missing data for initiation in the past 12 months (e.g., if the AFM for a specific prescription drug was answered as “don’t know”) without any other reports of initiation more than 12 months prior to the interview.

Respondents who reported that they initiated the misuse of one or more prescription drugs in a category more than 12 months before being interviewed (e.g., if a 20-year-old respondent reported first misusing OxyContin® at age 18) were skipped out of this follow-up question because their initiation data for misuse indicated that they initiated the misuse of some prescription drugs in that category more than 12 months prior to the interview date.

If respondents were asked this follow-up question, and they answered it as “no,” meaning that they had never misused any prescription drug in that category more than 12 months prior to the interview date, then it was concluded that the respondents initiated the misuse of all prescription drugs in that category in the past 12 months. However, if respondents answered the follow-up as “yes,” then they would *not* be considered recent initiates for the misuse of any prescription drug in that category (e.g., pain relievers) because this response meant that the respondents initiated the misuse of some other drug in the category more than 12 months ago.

The edited variables for indicating whether respondents initiated the misuse of all prescription drugs in a given category in the past 12 months were PNRNMINIT for pain relievers (corresponding to question PRL03), TRQNMINIT for tranquilizers (corresponding to TRL03), STMNMINIT for stimulants (corresponding to STL03), and SEDNMINIT for sedatives (corresponding to question SVL03). Creation of the edited variable PNRNMINIT for the initiation of misuse of prescription pain relievers in the past 12 months is described as an example.

The following edits were implemented for PNRNMINIT if respondents misused prescription pain relievers in the past 12 months:

- If respondents answered question PRL03 as “yes” or “no,” then PNRNMINIT was reverse coded.
 - If PRL03 was answered as “no” (PRL03 = 2), then PNRNMINIT was set to a value of 1, where 1 = Past year initiate.
 - If PRL03 was answered as “yes” (PRL03 = 1), then PNRNMINIT was set to a value of 2, where 2 = Not a past year initiate.
- If respondents answered question PRL03 as “don’t know” or “refused,” then PNRNMINIT retained the code for the answer from PRL03 (e.g., 94 = DON’T KNOW).
- If question PRL03 was skipped because respondents reported that they first misused any pain reliever more than 12 months prior to the interview, the PNRNMINIT was coded as 4, where 4 = Not a past year initiate LOGICALLY ASSIGNED (from skip pattern).

Additional edits when PRL03 had been skipped because respondents did not report the misuse of prescription pain relievers in the past 12 months took into account data from the recency of misuse variable PNRNMREC (Section 6.3.5).

- If respondents last misused any pain reliever more than 12 months ago (PNRNMREC = 3), then PNRNMINIT was set to a value of 93, where 93 = DID NOT MISUSE PAIN RELIEVERS IN THE PAST 12 MONTHS.

- If respondents did not misuse pain relievers in the past 12 months but it was unknown whether the respondents misused pain relievers more than 12 months ago or that they never misused pain relievers (PNRNMREC = 83; Sections 6.3.4.2 and 6.3.5), then PNRNMINIT was set to a value of 83, where 83 = NOT A PAST YEAR INITIATE (AT-RISK STATUS UNKNOWN).¹⁰⁷
- If respondents misused pain relievers in their lifetime but their most recent period of misuse was unknown (PNRNMREC = 9), then PNRNMINIT was coded as 98 (i.e., blank). PNRNMINIT retained a missing value for these respondents because they were potential misusers in the past 12 months (and potential initiates in that period).
- If PNRNMREC had been coded as 98, then PNRNMINIT also was coded as 98.

6.3.6.2 Age, Year, and Month of Initiation for Individuals Who Initiated Misuse in the Past 12 Months

For each specific prescription drug that respondents misused in the past 12 months, respondents were asked to report their age when they first misused that drug (AFM). If respondents reported that they first misused a drug (e.g., Vicodin®) within 1 year of their current age, they were asked to report the year and month when they first misused that drug (YFM and MFM, respectively). Recoded variables for the age, year, and month of first misuse also were created for the overall prescription drug categories (e.g., pain relievers) for respondents who initiated the misuse of all prescription drugs in that category in the past 12 months.

6.3.6.2.1 Age-, Year-, and Month-of-Initiation Variables for Specific Prescription Drugs

Unlike the logic in the other substance use sections, a “hard error” rather than a consistency check was triggered if respondents reported an AFM for a specific prescription drug that was greater than their current age. A hard error does not allow respondents to move to the next question until they enter an AFM that is consistent with their current age or they change the AFM answer to “don’t know” or “refused.” In contrast, the consistency checks in the other substance use sections give respondents a single opportunity to enter an AFU that is consistent with their current age but do not require respondents to resolve the inconsistency.

Because of the use of hard errors for prescription drugs when AFMs were inconsistent with respondents’ current ages, the AFM variables for specific prescription drugs would have values less than or equal to the current age or values for “don’t know” or “refused,” unless a different final age value had been assigned (see Section 4.2.2). In these situations, if the AFM was greater than the final age, the edited AFM was coded as bad data.

¹⁰⁷ Respondents who never misused prescription drugs in a given category in their lifetime are considered to be “at risk” for initiating misuse in the past 12 months. However, respondents who initiated the misuse of prescription drugs in a given category more than 12 months prior to being interviewed are no longer considered to be at risk for initiating misuse in the past 12 months because they already initiated misuse more than 12 months ago. If it is unknown whether respondents misused any prescription drug in a category in their lifetime, then their status for being at risk for initiation also is unknown.

Similar to the logic in other substance use sections, the CAI program compared the AFM for a given prescription drug with the age that was calculated from the YFM, MFM, and date of birth (subsequently referred to as the MYR1ST age, such as MYR1STPR1, for Vicodin®). A consistency check was triggered if the calculated MYR1ST age for that prescription drug differed from the corresponding AFM. Therefore, the edits described in [Table B.6](#) also applied to the AFM, YFM, and MFM variables for specific prescription drugs.

If the AFM for a specific prescription drug was 2 or more years younger than the respondent's current age, then the edited YFM and MFM were assigned legitimate skip codes (Section 2.4.2). Specifically, if the YFM and MFM had been skipped, then they were coded as 9999 and 99, respectively. If respondents initially reported first misuse within 1 year of their current age but changed the AFM in response to a consistency check to being 2 or more years younger than the current age ([Table B.6](#)), then any values in the YFM and MFM were overwritten with codes of 9989 and 89, respectively.

Additional edits to the AFM, YFM, and MFM variables (e.g., VICONMAGE, VICONMYFU, and VICONMMFU for Vicodin®) reflected the values that had been assigned through editing to the variable corresponding to the lead question for misuse (e.g., VICONMYR).

- If the lead question for misuse had a value of 2 (No), 4 (No LOGICALLY ASSIGNED), or 93 (Did not use), then blank values in the AFM, YFM, and MFM were replaced with codes of 993, 9993, and 93, respectively. Nonblank values were overwritten with corresponding codes of 983, 9983, and 83.
- If the lead question for misuse had a value of 91 because respondents had not misused any prescription drugs in that category in their lifetime, then blank values in the AFM, YFM, and MFM were replaced with codes of 991, 9991, and 91, respectively. Nonblank values were overwritten with corresponding codes of 981, 9981, and 81.
- If the lead question for misuse had a value of 81 because of the edits that were described in Section 6.3.4.4, the AFM, YFM, and MFM were globally assigned codes of 981, 9981, and 81, respectively.
- Otherwise, if the lead question for misuse had a value of 98 for blank (e.g., if the question for misuse was not asked because respondents answered the screener question for any past year use as "don't know" or "refused"), the AFM, YFM, and MFM retained codes of 998, 9998, and 98, respectively.

6.3.6.2.2 Age-, Year-, and Month-of-Initiation Variables for Initiates in the Past 12 Months

Recoded AFM, YFM, and MFM variables also were created for the misuse of any prescription drug in a given psychotherapeutic category (e.g., pain relievers) for respondents who reported initiating the misuse of all prescription drugs in that category in the past 12 months. If initiates in the past 12 months reported the misuse of only one prescription drug in that category, then creation of the recoded AFM, YFM, and MFM variables was straightforward. The AFM, YFM, and MFM from the specific prescription drug were transferred to the corresponding recoded variables for initiation of any prescription drug in that category. The general principles listed below were followed for creating these recoded AFM, YFM, and MFM variables when recent

initiates reported the misuse of more than one prescription drug in the past 12 months and they had no missing initiation data for the specific prescription drugs.

- The minimum AFM was chosen as the AFM for the overall category. If this AFM was reported for only one prescription drug in that category, then the YFM and MFM from the specific prescription drug corresponding to that AFM were assigned to the YFM and MFM variables for the first misuse of any prescription drug in that category.
- If the minimum AFM was reported for more than one prescription drug, then the minimum YFM was chosen as the YFM for the overall category. If this YFM was reported for only one prescription drug in that category, then the MFM from the specific prescription drug corresponding to that YFM was assigned to the corresponding MFM variable for the first misuse of any prescription drug in that category.
- If the minimum AFM and YFM were reported for more than one prescription drug, then the minimum MFM was chosen as the MFM for the overall category.

Focusing on initiates of misuse in the past 12 months facilitated the creation of recoded variables in the following situations:

- By definition, first misuse for initiates in the past 12 months would have occurred at the respondent's current age or at the age that is 1 year earlier than the respondent's current age. Consequently, if recent initiates reported first misusing one prescription drug at the age that was 1 year earlier than their current age but they had missing AFM data for other prescription drugs, it could still be inferred that the AFM for any prescription drug in that category was 1 year earlier than the current age.
- First misuse for initiates in the past 12 months would have occurred in the current year or the year immediately prior to the current year. Consequently, if recent initiates in 2023 reported first misusing one prescription drug in the year immediately prior to the current year, but they had missing YFM data for other prescription drugs, it could still be inferred that the YFM for any prescription drug in that category was in the year immediately prior to the current year.

Additional edits for the AFM, YFM, and MFM when respondents did not report initiation in the past 12 months took into account data from the corresponding initiation variable (e.g., PNRNMINIT). Edits for PNRNMAGE, PNRNMYFU, and PNRNMMFU when PNRNMINIT did not indicate initiation of the misuse of prescription pain relievers in the past 12 months are described in the examples below.

- If PNRNMINT had a value of 2 or 4 because respondents misused in the past 12 months but initiated misuse more than 12 months ago, then PNRNMAGE, PNRNMYFU, and PNRNMMFU were assigned codes of 993, 9993, and 93, respectively.
- If PNRNMINT was coded as 93 because respondents did not misuse pain relievers in the past 12 months, then PNRNMAGE, PNRNMYFU, and PNRNMMFU also were assigned codes of 993, 9993, and 93, respectively.

- If PNRNMINT was coded as 83 because respondents did not misuse pain relievers in the past 12 months but their lifetime misuse status was unknown, then PNRNMAGE, PNRNMYFU, and PNRNMMFU were assigned codes of 983, 9983, and 83, respectively.
- If PNRNMINT was coded as 91 because respondents did not misuse pain relievers in their lifetime, then PNRNMAGE, PNRNMYFU, and PNRNMMFU were assigned codes of 991, 9991, and 91, respectively.
- Otherwise, if PNRNMINIT had a missing value, then PNRNMAGE, PNRNMYFU, and PNRNMMFU were assigned codes for blank (i.e., 998, 9998, and 98, respectively). For example, if PNRNMINIT = 94 (i.e., don't know), then the AFM, YFM, and MFM variables retained missing values.

6.3.7 Misuse of Prescription Drugs in the Past 30 Days

A key component of editing the variables for misuse of prescription drugs in the past 30 days involved comparing these variables with the corresponding recency variable for misuse (e.g., PNRNMREC for pain relievers). If the question pertaining to misuse of any prescription drug in the category in the past 30 days (e.g., PRM01 for pain relievers; edited variable PNRNM30D) had been answered as "no" (e.g., PNRNM30D = 2), then the code of 2 for "no" was retained in the edited variable to indicate that the respondent did not misuse any prescription drug in that category in the past 30 days.

- If the recency had a value of 2 or 3 to indicate that the most recent misuse definitely occurred more than 30 days prior to the interview, then the skipped variables pertaining to misuse in the past 30 days were assigned a code of 93 (DID NOT MISUSE IN THE PAST 30 DAYS).
- If the recency had been assigned a code of 83 because the only misuse in the past 12 months involved the misuse of OTC drugs (Sections 6.3.4.2 and 6.3.5), then any responses in the 30-day questions were overwritten with a code of 83 in the edited variables (DID NOT MISUSE IN THE PAST 30 DAYS Logically assigned). If questions had been skipped (e.g., if the lead question for any misuse had been answered as "no"), then the edited variables were assigned a code of 93.
- If the recency had a value of 91 to indicate that respondents had never misused (or used) any prescription drug in that category in their lifetime, then the skipped variables for misuse in the past 30 days were assigned a code of 91 (NEVER USED/MISUSED).
- If the recency had a value of 8 or 9, then skipped variables for misuse in the past 30 days typically retained a code of 98 (blank). An exception for the misuse of prescription drugs with alcohol in the past 30 days is discussed in Section 6.3.7.2.
- Otherwise, if the recency had a value of 98 (i.e., blank), then the variables for misuse in the past 30 days were assigned a missing value code to indicate that lifetime misuse or no lifetime misuse was unknown. If the variables for misuse in the past 30 days had a code of 98 (blank), then these variables retained that code of 98. Otherwise, values were overwritten with a code of 85 (BAD DATA Logically assigned).

If respondents reported that they initiated the misuse of any specific prescription drugs in the 30 days prior to the interview,¹⁰⁸ then they were skipped out of the question that asked whether they misused any prescription drug in that category in the past 30 days (e.g., PRM01 for pain relievers). Therefore, it was not possible for respondents to report that they initiated the misuse of prescription drugs in a given category in the past 30 days and then to answer the question for misuse in that period as “no.” When the question for misuse of any prescription drug in a category had been skipped because respondents reported that they initiated the misuse of specific prescription drugs in the past 30 days, then the edited variable (e.g., PNRNM30D) was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

When respondents reported that they initiated misuse of specific prescription drugs in the past 30 days, they were eligible to be asked the follow-up questions pertaining to misuse in the past 30 days (i.e., frequency of misuse, misuse with alcohol). Consequently, data would exist in the additional edited variables for misuse in the past 30 days when the lead question for any misuse in the past 30 days had been assigned a code of 5.

Remaining edits that are described in this section pertain to respondents who misused any prescription drug in a given category within the past 30 days.

6.3.7.1 Frequency of Misuse of Prescription Drugs in the Past 30 Days

If respondents misused prescription drugs in a given category in the past 30 days, they were asked to report the number of days in that period on which they misused any prescription drug in that category (e.g., question PRM02 for pain relievers). Thus, the edited variables for the frequency of misuse in the past 30 days (e.g., PNRNM30FQ, corresponding to question PRM02) were continuous variables, with valid values ranging from 1 to 30.

If respondents did not know or refused to report the exact number of days on which they misused any prescription drug in a category in the past 30 days, they were asked in a follow-up question (e.g., PRM02DKRE for pain relievers) to give their best estimate of the number of days they misused prescription drugs in that category in the past 30 days. Similar to the edited variables for the estimated 30-day frequency in other sections of the interview (Section 6.2.4.4), these edited variables (e.g., PNRNM30ES, corresponding to question PRM02DKRE) were categorical, based on the following response options: (1) 1 or 2 days, (2) 3 to 5 days, (3) 6 to 9 days, (4) 10 to 19 days, (5) 20 to 29 days, or (6) all 30 days. If a value between 1 and 30 days was reported in the 30-day frequency question (e.g., PRM02), then the edited variable for the estimated 30-day frequency was assigned a code of 99 (LEGITIMATE SKIP).

Respondents could not report that they misused prescription drugs on “0 days” in the past 30 days, because respondents had been asked a “yes/no” question for whether they misused prescription drugs in that category in the past 30 days, or else they reported that they had

¹⁰⁸ Respondents were identified as having initiated the misuse of a prescription drug in the past 30 days if they (1) reported that they first misused that drug in the same month that they were interviewed; or (2) first misused at their current age (i.e., if not in the same month as the interview month), and their birthday was within 30 days of the interview date.

initiated misuse of specific prescription drugs in the past 30 days. Especially for respondents who were directly asked whether they misused any prescription drugs in a category in the past 30 days (e.g., PRM01 for pain relievers), an explicit answer of “yes” would indicate that respondents had misused prescription drugs in that category on at least 1 day in that period. Because the 30-day frequency questions for prescription drugs did not allow respondents to report misuse on 0 days in that period, the edits that were described in [Table 6.2](#) for the use of substances on 0 days in the past 30 days did not apply to the editing of the 30-day frequency variables for prescription drugs.

However, if respondents reported for all specific prescription drugs in a category that they misused in the past 12 months that they initiated misuse within 30 days of the interview date (regardless of whether they were past year initiates for the misuse of any prescription drug in that category¹⁰⁹), then it was possible for them to report a frequency of misuse that was greater than the number of days they could have misused based on their initiation data. For example, suppose respondents were not initiates in the past 12 months for the misuse of prescription pain relievers and the only prescription drug they misused in the past 12 months was OxyContin®. If these respondents reported first misusing OxyContin® in July 2023 and were interviewed on July 15, 2023, then the maximum number of days they could have misused OxyContin® in the past 30 days since their first misuse would be 15 days. This situation also held if respondents were initiates in the past 12 months, misused only OxyContin® in that period, and initiated misuse within 15 days of their interview date. For both scenarios, a report of misuse of prescription pain relievers on more than 15 days in the past 30 days would be inconsistent with the respondents’ data for the first misuse of OxyContin®. When these responses occurred, the 30-day frequency (e.g., PNRNM30FQ) was “trimmed” back to the value that would be consistent with the maximum number of days that these respondents could have misused prescription drugs in that category in the past 30 days. In the example cited in this paragraph, PNRNM30FQ would be trimmed back to a value of 15.

6.3.7.2 Misuse of Prescription Drugs with Alcohol in the Past 30 Days

Respondents who reported that they misused any prescription drug in a given category in the past 30 days and who also reported previously in the alcohol section that they last used alcohol in that period were asked whether they misused prescription drugs in that category in the past 30 days while they were drinking alcohol or within a couple of hours of drinking (e.g., question PRM03 for pain relievers; edited variable PNRNM30AL). In addition to respondents not being asked this question if they did not misuse a particular category of prescription drugs in the past 30 days, respondents were skipped out of this question if they did not report alcohol use in the past 30 days. If respondents reported misuse in the past 30 days for a given category of prescription drugs but the alcohol recency variable ALCREC indicated unambiguously that the most recent use of alcohol occurred more than 30 days ago (i.e., ALCREC = 2 or 3), then the edited variable for misuse with alcohol in the past 30 days (e.g., PNRNM30AL) was assigned a

¹⁰⁹ As noted in Section 6.3.6.1, for example, respondents could report that they were past year initiates of misuse for all specific prescription drugs they misused in the past 12 months but report that they misused other prescription drugs more than 12 months ago (e.g., if question PRL03 for pain relievers was answered as “yes”).

code of 99 (LEGITIMATE SKIP). Similarly, if respondents never used alcohol but they misused prescription drugs in a given category in the past 30 days, then the edited variable for misuse with alcohol in the past 30 days was assigned a code of 99.

However, if respondents indicated prescription drug misuse in the past 30 days, the 30-day frequency was skipped, and it was unknown whether respondents misused alcohol in the past 30 days, then the 30-day variable for misuse with alcohol retained a code of 98 (blank). For example, if ALCREC had been coded as 9 because respondents last used alcohol at some point in their lifetime (and potentially in the past 30 days) and respondents misused pain relievers in the past 30 days (i.e., PNRNMREC = 1), then PNRNM30AL retained a code of 98 if the corresponding question (e.g., PRM03) had been skipped.

In the assignment of codes when respondents legitimately skipped out of a question for the misuse of prescription drugs with alcohol in the past 30 days, the edits described at the beginning of Section 6.3.7 when respondents did not misuse prescription drugs in the past 30 days took precedence over situations in which respondents did not use alcohol in the past 30 days. Thus, for example, if ALCREC had been coded as 9 (last used alcohol at some point in the lifetime) but respondents last misused prescription drugs in a given category more than 30 days ago but within the past 12 months (e.g., PNRNMREC = 2 for pain relievers), then the edited variable for misuse with alcohol (e.g., PNRNM30AL) could be assigned a code of 93 despite it not being known whether the respondents used alcohol in the past 30 days.

Also, respondents could originally report that they last used alcohol in the past 30 days but report that they used alcohol on 0 days in that period. As noted in [Table 6.2](#), ALCREC was coded as 8 (Used at some point in the past 12 months) when this occurred. Respondents who confirmed in the alcohol section of the interview that they used alcohol on 0 days in the past 30 days were not asked questions about their misuse with alcohol in the past 30 days (PRM03, TRM03, STM03, and SVM03).

6.3.8 Ways of Misuse in the Past 12 Months

Respondents who misused one or more prescription drugs in a given category in the past 12 months were asked to report all of the ways in which they misused these prescription drugs in the past 12 months (e.g., question PRYWAYS for pain relievers). These questions applied to all prescription drugs that respondents misused in the past 12 months and at any point in that period. As an example, response options in PRYWAYS for pain relievers were (1) I used prescription pain relievers without a prescription of my own, (2) I used prescription pain relievers in greater amounts than they were prescribed, (3) I used prescription pain relievers more often than they were prescribed, (4) I used prescription pain relievers for longer than they were prescribed, and (5) I used prescription pain relievers in some other way a doctor did not direct me to use them.

These questions for the ways in which respondents misused prescription drugs in the past 12 months were “enter all that apply” questions. Therefore, separate variables were created that corresponded to each response option. The edited variables for the response options in PRYWAYS were PNRWYNORX (use without a prescription), PNRWYGAMT (use in greater

amounts than directed), PNRWYOFTN (use more often than directed), PNRWYLNGR (use longer than directed), and PNRWYOTWY (use in some other way not directed). As noted in Section 2.4.4, these variables were coded as 1 (Yes) if a particular way of misuse was reported from the list or as 2 (No [not entered]) if a particular way of misuse was not chosen but at least one of the other ways of misuse was chosen from the list.

A key component of editing the variables for the ways in which respondents misused prescription drugs in the past 12 months involved comparing these variables with the corresponding recency variable for misuse (e.g., PNRNMREC for pain relievers).

- If the recency had a value of 3 to indicate that the most recent misuse occurred more than 12 months prior to the interview, then the skipped variables pertaining to the ways of misuse in the past 12 months were assigned a code of 93 (DID NOT MISUSE IN THE PAST 12 MONTHS).
- If the recency had been assigned a code of 83 because the only misuse in the past 12 months involved the misuse of OTC drugs (Sections 6.3.4.2 and 6.3.5), then any responses in the questions for the ways of misuse in the past 12 months were overwritten with a code of 83 in the edited variables (DID NOT MISUSE IN THE PAST 12 MONTHS Logically assigned). If certain ways of misuse had not been chosen from the list, then the corresponding edited variables were assigned a code of 93.
- If the recency had a value of 91 to indicate that respondents had never misused (or used) any prescription drug in that category in their lifetime, then the skipped variables for the ways of misuse in the past 12 months were assigned a code of 91 (NEVER USED/MISUSED).
- If the recency had a value of 9, then skipped variables for misuse in the past 12 months retained a code of 98 (blank).
- Otherwise, if the recency had a value of 98 (i.e., blank), then the variables for the ways of misuse in the past 12 months were assigned a missing value code to indicate that lifetime misuse or no lifetime misuse was unknown. If the variables for the ways of misuse had a code of 98 (blank), then these variables retained that code of 98. Nonblank values were overwritten with a code of 85 (BAD DATA Logically assigned).

In addition, respondents could report that the only way that they misused prescription drugs in a given category in the past 12 months was without a prescription of their own. However, respondents also could report that they got their last prescription drug through a prescription, either from one doctor or from more than one doctor. Although it would be inconsistent for respondents to have misused prescription drugs only without a prescription of their own in the past 12 months and to have gotten the last prescription drug that they misused through a prescription they had, no editing was done to the data for the ways of misuse. Instead, a flag variable was set (e.g., PNRNORXFG for pain relievers) to alert analysts when this pattern occurred.

6.3.9 Injection of Prescription Stimulants in the Past 12 Months or Past 30 Days

Respondents who reported that they misused prescription stimulants at any point in the past 12 months were asked whether they used a needle to inject prescription stimulants at any time in this period (question STYNDL). Respondents who answered STYNDL as “yes” and who reported that they misused stimulants in the past 30 days were asked in question STYNDL30 to report when they last used a needle to inject stimulants (i.e., within the past 30 days or more than 30 days ago but within the past 12 months). Logically, respondents who misused all stimulants more than 30 days ago but within the past 12 months would not have used a needle to inject stimulants in the past 30 days.

The edited variables corresponding to these questions were STMNDLYR (for question STYNDL) and STMNDLREC (for STYNDL30). STMNDLYR and STMNDLREC were edited according to the recency of misuse of prescription stimulants (STMNMREC).

- The following edits were implemented if respondents misused stimulants in the past 30 days (STMNMREC = 1):
 - If STYNDL was answered as “yes” and STYNDL30 indicated misuse with a needle in the past 30 days (STNDL30 = 1), then both STMNDLYR and STMNDLREC were coded as 1. A code of 1 meant “Yes” for STMNDLYR and “Within the past 30 days” for STMNDLREC.
 - If STYNDL was answered as “yes” and STYNDL30 indicated misuse with a needle more than 30 days ago but within the past 12 months (STNDL30 = 2), then STMNDLYR was coded as 1 (Yes). STMNDLREC was assigned a code of 2 (More than 30 days ago but within the past 12 months).
 - If STYNDL was answered as “no,” then STMNDLYR was coded as 2 (No), and STMNDLREC was assigned a code of 99 (LEGITIMATE SKIP).
- The following edits were implemented if respondents last misused stimulants more than 30 days ago but within the past 12 months (STMNMREC = 2):
 - If STYNDL was answered as “yes,” then STMNDLYR was coded as 1. Because STYNDL30 had been skipped, STMNDLREC was assigned a code of 12, meaning: More than 30 days ago but in the past 12 months (STMNMREC=2).
 - If STYNDL was answered as “no,” then STMNDLYR was coded as 2, and STMNDLREC was assigned a code of 99.
 - If STYNDL was answered as “don’t know” or “refused,” then STMNDLYR retained the relevant missing data code (e.g., 97 for “refused”), and STMNDLREC retained a code of 98 (blank).

- When respondents were logically assigned to have last misused stimulants at some point in the past 12 months (STMNMREC = 8), it was not known whether they misused stimulants in the past 30 days. Consequently, they were skipped out of question STYNDL30 for the last time that they injected stimulants. The following edits were implemented when STMNMREC was coded as 8:
 - If STMNDLYR was coded as 2 because STYNDL was answered as “no,” then STMNDLREC was assigned a code of 99. Even if these respondents potentially misused stimulants in the past month, their reports that they did not inject stimulants in the past 12 months meant that the question about their most recent injection of stimulants did not apply.
 - If STYNDL was answered as “yes” and STMNDLYR was coded as 1, then STMNDLREC retained a code of 98 (blank), because it was unknown whether respondents last misused stimulants in the past 30 days or last misused them more than 30 days ago but within the past 12 months.
 - If STYNDL was answered as “don’t know” or “refused” and STMNDLYR retained the corresponding missing data code, then STMNDLREC retained a code of 98 (blank).
- If respondents last misused stimulants more than 12 months ago (STMNMREC = 3), then STMNDLYR and STMNDLREC were both assigned a code of 93 (DID NOT MISUSE IN THE PAST 12 MONTHS).
- If STMNMREC had been assigned a code of 83 because the only misuse in the past 12 months involved the misuse of OTC drugs (Sections 6.3.4.2 and 6.3.5), then any responses to STYNDL and STYNDL30 were overwritten with a code of 83 in the edited variables STMNDLYR and STMNDLREC, respectively (DID NOT MISUSE IN THE PAST 12 MONTHS Logically assigned). If STYNDL and STYNDL30 had been skipped, then the edited variables were assigned a code of 93.
- If STMNMREC had a value of 91 to indicate that respondents had never misused (or used) prescription stimulants in their lifetime, then STMNDLYR and STMNDLREC were assigned a code of 91 (NEVER USED/MISUSED).
- If STMNMREC had a value of 9 and the stimulant needle use variables had been skipped, then STMNDLYR and STMNDLREC retained a code of 98 (blank).
- Otherwise, STMNDLYR and STMNDLREC were assigned a code of 98 (blank) to indicate that lifetime misuse or no lifetime misuse was unknown.

6.3.10 Identification of the Last Prescription Drug Misused in the Past 12 Months

Questions about respondents’ reasons for misuse and how they obtained prescription drugs pertained to the last prescription drug in a given category that respondents misused in the past 12 months. For example, if respondents who misused prescription pain relievers in the past 12 months reported that they misused Vicodin® the last time, then question PRYMOTIV asked the respondents to think about the last time that they misused Vicodin® in the past 12 months and to report the reasons for misusing Vicodin® the last time.

If respondents reported that they misused more than one prescription drug in a category in the past 12 months, they were asked to report which drug they misused the last time (e.g., question PRYLAST for pain relievers). In these questions, the response options for the prescription drugs that respondents misused in the past 12 months were filled in sequentially according to the drugs that respondents reported misusing. Consequently, the response options could have different meanings according to the specific prescription drugs that respondents misused. For example, if a respondent misused Vicodin®, generic hydrocodone, and OxyContin® in the past 12 months, then response option 1 in PRYLAST was filled with "Vicodin," response option 2 was filled with "Hydrocodone (generic)," and response option 3 was filled with "OxyContin." If a second respondent misused generic hydrocodone, OxyContin®, and generic morphine in the past 12 months, then response option 1 in PRYLAST was filled with "Hydrocodone (generic)," response option 2 was filled with "OxyContin®," and response option 3 was filled with "Morphine (generic)."

In order for variables pertaining to the last prescription drug that was misused in the past 12 months to have one, and only one, meaning for all respondents (e.g., category 1 for pain relievers always pertaining to Vicodin® as the last pain reliever that was misused), corresponding recoded variables for the last prescription drug were created as part of the CAI program (e.g., RPRYLAST, corresponding to PRYLAST). For example, if "Hydrocodone (generic)" had been reported in PRYLAST as the last prescription pain reliever that a respondent misused (regardless of the drug's order in the list), then RPRYLAST was coded as 5 because generic hydrocodone was the fifth pain reliever in the questionnaire for pain relievers.

The recoding in the CAI program also took into account situations in which respondents reported the misuse of only one prescription drug in that category in the past 12 months. For these respondents, the CAI program automatically filled in the name of that prescription drug in the questions about the reasons for misuse and how respondents obtained the last prescription drug that they misused. For example, if respondents reported that the only pain reliever they misused in the past 12 months was generic hydrocodone, then respondents were asked to report their reasons for misusing generic hydrocodone and how they obtained it the last time, and RPRYLAST was coded as 5 (Hydrocodone [generic]). These recoded variables that were created from the CAI program were used as the starting point for creating the edited variables for the last prescription drug that was misused in the past 12 months.

Minimal additional editing was done to these variables. However, if respondents did not misuse prescription drugs in a given category in the past 12 months, then the edits that were described previously in Section 6.3.8 were implemented. For example, if respondents last misused prescription pain relievers more than 12 months ago (PNRNMREC = 3), then PNRNMLAS1 was assigned a code of 93.

6.3.11 Reasons for the Last Misuse of Prescription Drugs in the Past 12 Months

As noted in Section 6.3.10, respondents were asked to report the reasons for their misuse of the last prescription drug in a category in the past 12 months (e.g., question PRYMOTIV for pain relievers). Respondents who reported "some other reason" as a reason for their last

misuse were asked to specify this reason (e.g. question PRYMOTOT for pain relievers). If respondents reported more than one reason, they were asked to report which one of these reasons was the main reason for their misuse (e.g., PRYMOT1 for pain relievers).

The response options for the reasons for misuse varied by prescription psychotherapeutic category. Response options for pain relievers were (1) to relieve physical pain, (2) to relax or relieve tension, (3) to experiment or to see what the drug is like, (4) to feel good or get high, (5) to help with my sleep, (6) to help me with my feelings or emotions, (7) to increase or decrease the effect(s) of some other drug, (8) because the respondent is “hooked” or has to have the drug, or (9) for some other reason. The same response options were presented for tranquilizers and sedatives, except that “to relieve physical pain” was not presented as an option. Instead, the first response option for both of these psychotherapeutic categories was “to relax or relieve tension.” Finally, the response options for stimulants were (1) to help lose weight, (2) to help concentrate, (3) to help be alert or stay awake, (4) to help study, (5) to experiment or to see what the drug is like, (6) to feel good or get high, (7) to increase or decrease the effect(s) of some other drug, (8) because the respondent is “hooked” or has to have the drug, or (9) for some other reason.

Although the response options varied by psychotherapeutic category, a common set of codes was used for coding the “OTHER, Specify” data for other reasons for misuse. [Table 6.5](#) lists the codes for “OTHER, Specify” data that corresponded to the response options that were presented to respondents for reasons for misuse. The use of a common set of codes for the reasons for misuse facilitated the use of the computer-assisted procedures described in Section 2.3.1 for the coding of “OTHER, Specify” data. For example, if respondents specified relief of physical pain as a reason for misuse, this response was coded as 1 regardless of whether they specified it as a reason for the misuse of pain relievers or as some other reason for the misuse of tranquilizers.

Table 6.5 “OTHER, Specify” Codes for Reasons for Misuse That Corresponded to Available Response Options

Code	Description	Comments
1	To relieve physical pain	Response option 1 for pain relievers. Not available as a precoded response option for the other prescription drugs.
2	To relax or relieve tension	Response option 2 for pain relievers and response option 1 for tranquilizers and sedatives. Not available as a precoded response option for stimulants.
3	To experiment or to see what it’s/they’re like	Response option 3 for pain relievers, response option 2 for tranquilizers and sedatives, and response option 5 for stimulants.
4	To feel good or get high	Response option 4 for pain relievers, response option 3 for tranquilizers and sedatives, and response option 6 for stimulants.
5	To help with my sleep	Response option 5 for pain relievers and response option 4 for tranquilizers and sedatives. Not available as a precoded response option for stimulants.
6	To help me with my feelings or emotions	Response option 6 for pain relievers and response option 5 for tranquilizers and sedatives. Not available as a precoded response option for stimulants.

Table 6.5 “OTHER, Specify” Codes for Reasons for Misuse That Corresponded to Available Response Options (continued)

Code	Description	Comments
7	To increase or decrease the effect(s) of some other drug	Response option 7 for pain relievers and stimulants and response option 6 for tranquilizers and sedatives.
8	Because I am “hooked” or I have to have it/them	Response option 8 for pain relievers and stimulants and response option 7 for tranquilizers and sedatives.
9	To help me lose weight	Response option 1 for stimulants. Not available as a precoded response option for the other prescription drugs.
10	To help me concentrate	Response option 2 for stimulants. Not available as a precoded response option for the other prescription drugs.
11	To help me be alert or stay awake	Response option 3 for stimulants. Not available as a precoded response option for the other prescription drugs.
12	To help me study	Response option 4 for stimulants. Not available as a precoded response option for the other prescription drugs.
13	I used it/them for some other reason	Response option 9 for pain relievers and stimulants and response option 8 for tranquilizers and sedatives. However, this was a “placeholder” code that was not used in the actual coding of “OTHER, Specify” data. Instead, new codes for other reasons started with level 14.

A key component of editing the variables for the reasons for the last misuse of prescription drugs in a given category in the past 12 months involved comparing these variables with the corresponding recency variable for misuse (e.g., PNRNMREC for pain relievers). If respondents did not misuse prescription drugs in a given category in the past 12 months, then the edits that were described previously in Section 6.3.8 were implemented. For example, if respondents last misused prescription pain relievers more than 12 months ago (PNRNMREC = 3), then the edited variables for the reasons for that last misuse of pain relievers were assigned codes of 93 (or 993 for the “OTHER, Specify” variable PNRRSOTRS for other reasons for misuse of pain relievers). Remaining edits that are described in this section applied to respondents who last misused prescription drugs in a category in the past 12 months.

Because these questions for the reasons for misuse of the last prescription drug in a given category in the past 12 months were “enter all that apply” questions, separate variables were created for each psychotherapeutic category that corresponded to each response option. As noted in Section 2.4.4, these variables were coded as 1 (Yes) if a particular reason for misuse was reported from the list, or as 2 (No [not entered]) if a particular reason for misuse was not chosen but at least one of the other reasons for misuse was chosen from the list. If respondents did not choose “some other reason” from the list, then the edited variable corresponding to the “OTHER, Specify” response (e.g., PNRRSOTRS for pain relievers) was assigned a code of 999 (LEGITIMATE SKIP).

No editing of the variables for any reason for misuse was done if respondents reported multiple reasons for their misuse of the prescription drug that they misused in the past 12 months and

they did not report “some other reason” for misuse. However, “OTHER, Specify” data were used to edit the variables for any reason for misuse. For example, if respondents reported that they misused pain relievers for some other reason and this other reason was to relieve physical pain (i.e., PNRRSOTRS coded as 1) and this response was not chosen as a reason in PRYMOTIV, then the edited variable PNRRSPAIN for the misuse of pain relievers to relieve physical pain was assigned a code of 3 (Yes LOGICALLY ASSIGNED). Similarly, if respondents reported that their other reason for misusing tranquilizers was to relax or relieve tension (i.e., TRQRSOTRS coded as 2 for the edited “OTHER, Specify” variable) and this response was not chosen as a reason in TRYMOTIV, then the edited variable TRQRSRELX for the misuse of tranquilizers to relax or relieve tension was assigned a code of 3.

As noted previously, if respondents reported more than one reason for their last misuse, they were asked to report which reason was their main reason (e.g., PRYMOT1 for pain relievers). As with the question for the last prescription drug that was misused in the past 12 months, however, response options in a given question for the main reason for the last misuse were filled in sequentially according to the reasons that respondents gave in the previous question (e.g., PRYMOTIV for pain relievers). Consequently, the response options could have different meanings according to the specific reasons that respondents previously reported.

In order for variables pertaining to the main reason for the last misuse for a given prescription drug category to have one, and only one, meaning for all respondents (e.g., response option 5, the main reason for the misuse of pain relievers always pertaining to misuse to help with sleep), corresponding recoded variables for the main reasons for the last misuse were created as part of the CAI program (e.g., RPRYMOT1, corresponding to PRYMOT1). For example, if misuse to help with sleep had been reported in PRYMOT1 as the main reason for the last misuse of pain relievers, then RPRYMOT1 was coded as 5, regardless of that reason’s order in the list from PRYMOT1.

These recoded variables that were created from the CAI program were used as the starting point for creating the edited variables for the main reason for the last misuse. When respondents reported only one reason for their last misuse, then the edited variable for the main reason for the last misuse (e.g., PNRRSMAIN for pain relievers) was assigned a code of 99 (LEGITIMATE SKIP).

“OTHER, Specify” data for other reasons for the last misuse also were used to edit the data for the main reason for misuse. Editing of the variables for the reason for last misuse took into account the available response options for a given prescription drug category. For example, the first response option for the reasons for the last misuse of tranquilizers was “To relax or relieve tension.” Therefore, if respondents reported that the “other” reason for misusing the last tranquilizer in the past 12 months was to relax or relieve tension (TRQRSOTRS = 2) and they reported that the main reason for the last misuse was, “The other reason I reported,” then TRQRSMAIN was coded as 11 (To relax or relieve tension LOGICALLY ASSIGNED).

In addition, when respondents reported more than one reason for their last misuse and they reported that the other reason they reported was the most important reason for their last misuse, the “OTHER, Specify” data for this other reason could have a missing value (i.e., bad

data, “don’t know,” or “refused”). When this pattern occurred, the edited variable for the main reason for misuse was coded as bad data (e.g., PNRSMMAIN = 85 for the main reason for the last misuse of pain relievers) because respondents did not substantiate what this “most important” other reason was. For example, if respondents reported misusing pain relievers the last time for some other reason but then reported that they did not know what this other reason was, it would not make logical sense for respondents to report that this unknown reason was the most important reason for their last misuse.

6.3.12 Sources of Prescription Drugs for the Last Misuse in the Past 12 Months

For all of the prescription psychotherapeutic drug categories, respondents were given a list of eight potential sources of prescription medications indicating where they obtained the last prescription drug in a given category that they misused in the past 12 months. These sources included (1) prescriptions from one doctor, (2) prescriptions from more than one doctor, (3) thefts from medical facilities, (4) prescription drugs obtained from friends or relatives for free, (5) buying them from friends or relatives, (6) taking them from friends or relatives without asking, (7) purchases from a drug dealer or other stranger, or (8) “in some other way.” Respondents who indicated that they obtained their last prescription drug in some other way were asked to specify what that other way was.

Respondents who reported that they obtained their last prescription drug in a given category from a friend or relative for free were asked a follow-up question (e.g., PRYFRL for pain relievers) to report how the friend or relative originally obtained the drug. Similar to their predecessor questions, respondents were given the same list of eight sources for where a friend or relative could have obtained the last prescription drug that the respondent misused. Respondents also had the option of indicating that the friend or relative obtained this prescription drug “in some other way.” In that situation, respondents then were asked to specify the other way that the friend or relative obtained the medication.

A key component of editing the variables for the source of the prescription drug in a given category that respondents last misused in the past 12 months involved comparing these variables with the corresponding recency variable for misuse (e.g., PNRNMREC for pain relievers). If respondents did not misuse prescription drugs in a given category in the past 12 months, then the edits that were described previously in Section 6.3.8 were implemented. For example, if respondents last misused prescription pain relievers more than 12 months ago (PNRNMREC = 3), then the edited variables for how they obtained their last pain reliever and how a friend or relative obtained the last pain reliever that respondents misused were assigned codes of 93. Remaining edits that are described in this section applied to respondents who last misused prescription drugs in a category in the past 12 months.

Respondents who misused prescription drugs in a given category in the past 12 months were asked to report the one way that they obtained the last prescription drug that they misused. If respondents got the last prescription drug they misused in more than one way, they were asked to report one of the ways as their best answer. Thus, for the edited variables PNRGOTLAS, TRQGOTLAS, STMGOTLAS, and SEDGOTLAS (pertaining to how respondents got the last

prescription pain relievers, tranquilizers, stimulants, or sedatives, respectively, the last time they misused them), codes in these variables corresponded to those in the corresponding questions PRYGOT, TRYGOT, STYGOT, and SVYGOT.

If respondents reported that they obtained the last prescription drug from a friend or relative for free in the past 12 months (e.g., PNRGOTLAS for pain relievers), similar edits applied to the variables pertaining to how friends or relatives obtained the medication that respondents last misused. Respondents could indicate only one way that the friend or relative obtained the prescription drug that respondents last misused.

Miscellaneous edits based on skip logic also applied to the source of psychotherapeutics variables. These edits applied to all of the prescription psychotherapeutic drugs, with situations being cited for pain relievers. For example, if respondents who misused prescription pain relievers in the past 12 months reported a way in which they obtained the last pain reliever they misused other than “in some other way,” then the corresponding “OTHER, Specify” variable (e.g., PNRGOTOTH, corresponding to question PRYGOTSP) was assigned a code of 99 (LEGITIMATE SKIP). In addition, if respondents reported a way in which they obtained the last pain reliever other than from a friend or relative for free, then the edited variables PNRFFGOT and PNRFFGTSP (corresponding to questions PRYFRL and PRYFRLSP, respectively) were assigned a code of 99.

Levels 1 through 7 in the “OTHER, Specify” variables were used for responses that corresponded to existing response options. For example, a code of 4 was assigned to PNRGOTOTH (corresponding to question PRYGOTSP) if respondents reported in PRYGOTSP that the other way they got the last pain reliever they misused in the past 12 months was from a friend or relative for free. Because it could be inferred that these respondents got the last pain reliever they misused from a friend or relative for free, PNRGOTLAS was assigned a code of 14 (Got from a friend or relative for free LOGICALLY ASSIGNED).

[Table B.10](#) describes additional edits that were relevant to the source of psychotherapeutics variables. For these edits, the variables for the source of pain relievers are used as examples, although these edits also applied to the other prescription psychotherapeutic drugs. As noted in Section 6.3.8, for example, a flag was set (e.g., PNRNORXFG for pain relievers) when respondents reported that they got their last prescription drug through a prescription (i.e., from one doctor or from more than one doctor) but the only way in which they misused prescription drugs in a given category in the past 12 months was without a prescription of their own. Where relevant, the edits that are described in [Table B.10](#) for how respondents obtained the last prescription drug they misused in the past 12 months also were applied to the variables for ways that friends or relatives obtained the last prescription drug that respondents misused.

6.3.13 Edit Rules for Prescription Drugs

With the redesign of the prescription drug questions in 2015, steps were taken toward better integrating editing and imputation procedures, focusing mainly on prescription drugs. These integration efforts involved the development of “edit rules” that directly identify patterns that are unallowable in the final analytic dataset. Edits described in this section are discussed in

more detail in sections specific to the respective drugs. The logic of these edit rules is as follows:

- Before the edited variables were finalized, code was run with the edited data to identify any unallowable inconsistencies. The final edited data were corrected if any unallowable inconsistencies were identified.
- The code that applies the edit rules was used to create bounds for 30-day frequency variables for hallucinogens, inhalants, methamphetamine, and the four prescription drug categories, as described in Section 6.5.2.4.5.1.
- The code that applies the edit rules was used in the imputation programs that assign the age at first misuse and date of first misuse for prescription drugs (Section 6.5.2.6.2.1) to create upper and lower bounds for the date of first misuse. This step simplified the code in part by making it unnecessary to segregate the past year initiates into missing data patterns before dates of first misuse were randomly assigned.
- Before the imputation-revised variables were finalized, a confirmation was run on the imputation-revised data to confirm that no unallowable inconsistencies remained in the data.

The advantages of these small modifications are twofold. First, a smaller dataset can be carried through the modified predictive mean neighborhood imputation programs because the estimated 30-day frequency, the birth date, the interview date, and the initiation data for misuse are no longer needed. Second, changes across the editing and imputation processes can be implemented more efficiently because common SAS® programs are shared across both processes. The sharing of common SAS® programs also reduces the potential for errors. [Table 6.6](#) lists the inconsistent patterns that could occur and their resolutions.

Table 6.6 Inconsistent Patterns in the Prescription Drug Modules and Their Resolution

Pattern	Inconsistent Pattern	Section References	Action
1	If a respondent (R) was known to have initiated misuse in the past month at the category level (e.g., for pain relievers), the 30-day frequency exceeded the number of days that the R could have misused in that period.	6.3.7.1 (Editing) 6.5.2.4.5.1 (Imputation)	The 30-day frequency was trimmed to the maximum number of days that the R could have misused.
2	The R is not a past year initiate, but for each specific drug that was misused in the past year, the R reported a month of first misuse (MFM) and year of first misuse (YFM) that was within 30 days of the interview date. As for Pattern 1, the 30-day frequency exceeded the number of days that the R could have misused in that period.	6.3.7.1 (Editing) 6.5.2.4.5.1 (Imputation)	The 30-day frequency was trimmed to the maximum number of days that the R could have misused.

Table 6.6 Inconsistent Patterns in the Prescription Drug Modules and Their Resolution (continued)

Pattern	Inconsistent Pattern	Section References	Action
3	For any specific prescription drug, the age at first misuse (AFM) was inconsistent with the YFM and MFM (i.e., the R failed to resolve the inconsistency when confronted with the consistency check).	6.3.6.2.1 and Table B.6 (Editing) 6.5.2.6.2.1 (Imputation)	The YFM/MFM or both the YFM/MFM and AFM were set to missing, depending on the answers in the consistency checks.
4	For any specific prescription drug, the AFM was inconsistent with the YFM and MFM only because the Rs' birthday was on the first day of the month and the MFM was in the same month as the birth month.	6.3.6.2.1 (Editing) 6.5.2.6.2.1 (Imputation)	The AFM reported by the R was replaced with an AFM that was 1 year older than the originally reported AFM.
5	The R reported only use without a prescription in the past 12 months (e.g., in PRYWAYS for pain relievers). However, the R also reported getting the last drug as a prescription from one doctor or from more than one doctor.	6.3.8 and Table B.10 (Editing) Not applicable for imputation	No editing was done, but a flag was set to alert analysts to this pattern.
6	The R reported more than one reason for misusing the last prescription drug and reported that the "other reason" was the most important reason for misuse. However, the "other" reason that was specified was coded with a missing value (i.e., bad data, "don't know," or "refused").	6.3.11 (Editing) Not applicable for imputation	The edited main reason variable was set to bad data because the R did not substantiate what this "most important" other reason was.
7	The R was interviewed on the last day of the month and was a past year initiate but reported first misuse in the same month as the interview in the previous calendar year.	Not applicable for editing 6.5.2.6.2.1 (Imputation)	The date of first misuse was imputed to be consistent with the initiation data, even though it is technically more than 1 year prior to the interview date. This handling is consistent with the 12-month date range presented to respondents in the interview.

6.4 Additional Edits Applied during Imputation Processing

Additional edits based on previously imputed variables were applied for selected variables during the imputation process and are discussed in this section. In general, these edits affected only a few records. Edits described in Sections 6.4.1 through 6.4.3 were implemented mostly to resolve residual inconsistencies that prevented the determination of a valid interval for the assignment of the date of first use (Sections 6.5.1.4.4 and 6.5.2.6.1). Edits described in Sections 6.4.4 and 6.4.5 were implemented to create base imputation variables for benzodiazepines and fentanyl product outcomes. These base imputation variables incorporate "OTHER, Specify" codes.

6.4.1 “Other” Hallucinogens

For respondents who were known to have never used “other” hallucinogens, certain logical deductions could be made regarding the relationship between data for the parent measure and data for the child measure if all the necessary conditions that also are described were met (see [Table 6.1](#)). Note that these edits also could have been applied to respondents who were imputed to lifetime nonuse of the “other” variable.

1. If the respondent was known never to have used “other” hallucinogens, the overall hallucinogens recency was past month, one of the child recencies was missing, and no other child recency was past month or missing, then the child recency that was missing was edited to past month. Similar deductions were made when the overall hallucinogens recency was past year not past month and lifetime not past year.
2. If the respondent was known never to have used “other” hallucinogens including ketamine, the tryptamine hallucinogens DMT/AMT/“Foxy,” and *Salvia divinorum*, only one child AFU value was missing, ¹¹⁰ and the minimum of the nonmissing child AFU values was greater than the parent AFU value, then the missing child AFU value was edited to the parent AFU value.
3. If the respondent was known never to have used “other” hallucinogens including ketamine, the tryptamine hallucinogens DMT/AMT/“Foxy,” and *Salvia divinorum*, only one child AFU value was missing (see [Table 6.1](#)), the minimum of the nonmissing child AFU values was equal to the parent AFU value, and the earliest of the nonmissing child months and years of first use was later than the parent MFU and YFU, then the missing child AFU value was edited to the parent AFU value. ¹¹¹

6.4.2 Respondents Imputed to Lifetime Use for Child Drug Variables

As discussed in Section 6.5, the first substance use variables to undergo imputation consisted of the lifetime substance use measures. The results of these imputations could restrict the range of plausible values for other substance use measures, and, therefore, based on this additional information obtained from the imputations, certain editing rules that were applied to the unedited recency and frequency data had to be reapplied. The list of these edits follows:

1. If the parent recency of use was known to be lifetime but not past year, and the respondent was imputed to lifetime use of the child drug(s), then the child recency was set to lifetime but not past year. This was done because the respondent could not have used the child drug more recently than the parent drug.
2. If the respondent used the parent drug on exactly 1 day in the past 12 months, the parent initiation data indicated a date of first use in the past year, and the respondent was imputed to lifetime use of the child drug, then the child recency of use was set equal to the parent recency of use, and the child 12-month frequency of use was set to

¹¹⁰ “Only one child AFU value was missing” refers to only LSD, PCP, and Ecstasy. AFU data are not available for other hallucinogen child drugs.

¹¹¹ This situation rarely occurs. Procedures flag this response pattern for further examination to ensure the automated edits are applied correctly.

1 day. This was done because the respondent could not have used the child drug on any days when the parent drug was not used, so the recencies and frequencies cannot differ. This edit only applied to crack, which is the only child drug with a 12-month frequency.

3. If the parent initiation data indicated a date of first use in the past year, the parent recency of use was past year but not past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year but not past month. This was done because the respondent could not have used the child drug more recently than the parent drug (eliminating the possibility of past month recency), and the respondent also could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency). This edit only applied to crack.
4. Similarly, if the parent initiation data indicated a date of first use in the past year, the parent recency of use was past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year (whether the respondent had used in the past month could not be determined). This was done because the respondent could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency). Again, this edit only applied to crack.

6.4.3 Age-at-First-Use Variables

The edits that are described in this section were implemented in preparation for the imputation process. These edits are designed to reduce processing time and to set the imputation indicators properly (i.e., to “logically assigned” instead of to “statistically imputed”). As with the other edits that were described previously, these edits affected only a small number of records.

1. If the parent AFU value was missing and the minimum of the child AFU values was 3 years, then the parent AFU value was edited to 3 years. This was done because respondents with AFU values of less than 3 years were ineligible to be donors (Sections 6.5.1.4.1 and 6.5.2.4.5.2).¹¹² This edit applied to all parent AFU variables: cigarettes, overall hallucinogens, and cocaine.
2. If the parent AFU was equal to the respondent’s current age, all missing child AFU values were edited to the same age. This edit applied to the AFU variables for daily cigarettes, and crack.¹¹³
3. If the parent AFU was equal to 1 less than the respondent’s current age, the child recency¹¹⁴ was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child AFU value was missing, then the child AFU value was assigned to 1

¹¹² This could be considered to be an imputation rule rather than an editing rule. Nevertheless, once the decision is made to prevent those with age-at-first-use values of less than 3 years from being donors, it is clear that these do not require any sort of stochastic imputation.

¹¹³ For LSD, PCP, and Ecstasy, this edit was accomplished through restrictions on the donor pool during the imputation step.

¹¹⁴ Because there was no recency question associated with daily cigarettes, the overall cigarette recency was used instead.

less than the respondent's current age. This was done because the child AFU cannot be less than AGE - 1, because the parent AFU is AGE - 1, and the respondent could not have begun using a child measure before using the parent measure. It also cannot be greater than AGE - 1, because the child recency of lifetime but not past year indicates that the respondents did not use the drug while at their current age (because they did not use the drug at all in the past year). If the respondents did not use the substance at all in the past year, then they could not have begun using the substance in the past year. Because the child AFU cannot be less than AGE - 1 or greater than AGE - 1, it must be equal to AGE - 1. This edit also applied to daily cigarettes and crack.

4. If the age at first cigarette use was equal to AGE - 3, cigarette recency was lifetime but not past 3 years, and age at first daily cigarette use was missing, then age at first daily cigarette use was assigned to AGE - 3. The logic is similar to the above edit: the age at first cigarette use precludes the possibility that the age at first daily cigarette use was less than AGE - 3, and the cigarette recency precludes the possibility that the age at first daily cigarette use was greater than AGE - 3.

6.4.4 Benzodiazepines Variables

Prior to imputing past year any use and past year misuse of benzodiazepines, base imputation variables were created. Respondents were classified as having used any benzodiazepine tranquilizer or sedative in the past year if they reported the use of one or more benzodiazepines: Alprazolam products (Xanax®, Xanax® XR, Alprazolam, Extended-Release Alprazolam), Lorazepam products (Ativan®, Lorazepam), Clonazepam products (Klonopin®, Clonazepam), Diazepam Products (Valium®, Diazepam), Flurazepam (also known as Dalmane®), Temazepam products (Restoril®, Temazepam), and Triazolam products (Halcion®, Triazolam). Additionally, respondents were classified as having used any benzodiazepine tranquilizer or sedative if they reported the misuse of another benzodiazepine tranquilizer as some "other tranquilizer" or the misuse of another benzodiazepine sedative as some "other sedative."

Respondents could specify a benzodiazepine prescribed as a tranquilizer (e.g., Xanax®) as some "other sedative" they misused in the past 12 months. Similarly, respondents could specify a benzodiazepine prescribed as a sedative (e.g., Halcion®) as some "other tranquilizer" they misused in that period. Regardless of whether specific benzodiazepines were reported as tranquilizers or sedatives, these responses were included in the aggregate measures of any use or misuse of benzodiazepines in the past 12 months because the benzodiazepine measures were constructed from data in both the tranquilizers and sedatives sections. Consistent with the principle of not editing across sections of the interview, however, reports of benzodiazepines in sections other than tranquilizers or sedatives were *not* included in the measures of benzodiazepine use or misuse.

6.4.5 Fentanyl Products Variables

Prior to imputing past year any use and past year misuse of fentanyl products, base imputation variables were created. Respondents were classified as having used fentanyl products if they reported the use of Duragesic®, Fentora®, or Fentanyl. Additionally, respondents were classified

as having used fentanyl products if they reported the misuse of another fentanyl product as some “other pain reliever.” These base imputation variables do not include illegally made fentanyl.

6.5 Imputation of Substance Use Variables

Two imputation methodologies were applied to the 2023 substance use data as described in Chapter 3. The first method, predictive mean neighborhood (PMN), was applied in a similar manner to previous years for a subset of the substance use variables. Substances imputed using PMN include cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine and crack, and heroin.

The second imputation method, modified predictive mean neighborhood (modPMN), was applied to the remaining substances including vaping of nicotine, CBD or hemp, inhalants, hallucinogens, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription sedatives, and prescription stimulants. Note that hallucinogen child drug lifetime measures were imputed using PMN methodology with the exception of the provisional imputation step, which used colocated stochastic imputation (CSI) procedures instead of hot-deck procedures. Although these measures appear as set 22 in [Table 6.7](#), they are not imputed until after overall hallucinogen lifetime has been imputed (set 23 in [Table 6.11](#)).

The substance use measures collected in the 2023 NSDUH included lifetime use, recency of any use, recency of use/misuse, frequency of use in the past 12 months, frequency of use/misuse in the past 30 days, past year initiation status, and AFU/AFM, MFU/MFM, and YFU/YFM. However, depending on the substance in question, only a subset of these measures was collected and imputed.

6.5.1 Variables Imputed Using PMN Methodology

[Table 6.7](#) summarizes the substances and substance use measures that were imputed using the PMN methodology. This table also indicates how these measures were segregated into units referred to as imputation sets. See Section 3.4 for more information on imputation sets.

Because there are numerous sets, some of the set-specific descriptions are explained as deviations from the procedures applied to an earlier set. For example, imputation set 2 is described in detail in Section 6.5.1.3, and the rest of the recency and frequency sets are described in Section 6.5.1.6 as deviations from the procedures described in Section 6.5.1.3.

Because substance use was highly correlated with age, and to facilitate more timely implementation of the imputation procedures, the model building and final assignment of imputed values for tobacco, alcohol, marijuana, cocaine and crack, and heroin use variables as well as hallucinogen child drug lifetime use variables were performed separately within three distinct age groups: 12 to 17, 18 to 25, and 26 or older.

Table 6.7 Substances and Substance Use Measures, PMN Imputation Sets

Substance	Substance Use Measures						Age at First Daily Use
	Lifetime Usage	Recency of Use	12-Month Frequency of Use	30-Day Frequency of Use	Ways of Use	Age at First Use	
Cigarettes	NM	Set 2 (12-Month Frequency N/A)			N/A	Set 3	Sets 4, 5 ¹
Smokeless Tobacco		Set 6 (12-Month Frequency N/A)			N/A	Set 7	
Cigars		Set 8 (12-Month Frequency N/A)			N/A	Set 9	
Pipes		Set 10 (12-Month and 30-Day Frequency N/A)			N/A	N/A	
Alcohol	Set 1		Set 11		N/A	Set 12	N/A
Marijuana			Set 13		Sets 14 and 15	Set 16	
Cocaine and Crack			Set 17		N/A	Set 18	
Heroin			Set 19		N/A	Set 20	
Binge Alcohol Use	N/A	N/A	N/A	Set 21	N/A	N/A	
Hallucinogen Child Drugs ²	Set 22	DM	N/A	N/A	N/A	DM	N/A

PMN = predictive mean neighborhood; DM = different method; N/A = not applicable; NM = never missing. Lifetime cigarette use is used to define a unit respondent and is therefore never missing.

¹ Prior to imputing age at first daily cigarette use, lifetime daily cigarette use must first be imputed.

² Hallucinogen child drugs lifetime use variables are imputed using PMN methodology, except that their provisional imputation steps use collocated stochastic imputation procedures. Although listed as set 22, imputation is not started until after the overall hallucinogen lifetime use has been imputed (set 23). Other measures for hallucinogen child drugs are imputed using modified predictive mean neighborhood (modPMN) methodology (indicated with DM). See [Table 6.11](#) for modPMN imputation sets.

6.5.1.1 Lifetime Substance Use (Imputation Set 1)

The lifetime substance use variables were imputed using the single response propensity (RP)/multiple prediction (PRD) type of PMN, as outlined in Section 3.4.3. In general, the response rates for lifetime substance use variables were very high, with less than 1 percent requiring imputation. These high response rates were observed, in part, because of the usable interview rule, which requires that a respondent answer “yes” or “no” to the question on lifetime use of cigarettes and that at least nine (9) of the following additional substances contain information about use or nonuse: (1) smokeless tobacco, (2) cigars, (3) alcohol, (4) marijuana, (5) cocaine (in any form), (6) heroin, (7) hallucinogens, (8) inhalants, (9) methamphetamine, (10) pain relievers, (11) tranquilizers, (12) stimulants (i.e., independent of methamphetamine), and (13) sedatives. See Section 2.2.3 for more information about the usable interview rule.

Because the single RP/multiple PRD type of PMN was used for the lifetime usage imputations, decisions had to be made on the order in which to fit the PRD models. Substances later in the sequence would have more covariates in their models, because substances earlier in the

sequence were used as covariates after provisional imputation. With the exception of lifetime cigarette use, the order in which the lifetime indicators of use in set 1 were imputed is shown in [Table 6.7](#).¹¹⁵ The lifetime use or nonuse of cigarettes was used to define a unit respondent for the NSDUH and, therefore, did not contain any missing values.

6.5.1.1.1 Response Propensity Step

The input to the weight adjustment model in the RP step for imputation set 1 was the preliminary analysis weight. The 2023 survey implemented automatic routing of the respondent through the questionnaire based on the respondent's answers, thereby skipping over (i.e., not asking the respondents) specific questions. Within each substance use section included in imputation set 1, one question was asked in order to establish whether respondents had ever used the substance in question during their lifetime. For more information on gate questions, see Section 6.2.1.

For individuals to be considered lifetime use item respondents, they must have complete data for all of the following substance use section gate questions: cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana, cocaine, crack, and heroin. See Appendix D for information on the covariates used in the RP models for these variables.

6.5.1.1.2 First Prediction Step (Lifetime Smokeless Tobacco Use)

The probability of lifetime smokeless tobacco use was modeled for item respondents within each age group, using the nonresponse-adjusted weights. SUDAAN®'s RLOGIST procedure was used to perform dichotomous logistic regression¹¹⁶ to determine the parameter estimates and probability of use for both respondents and nonrespondents.

6.5.1.1.3 First Provisional Hot-Deck Step (Lifetime Smokeless Tobacco Use)

To use lifetime usage of a given substance as a covariate for a substance later in the sequence, it was necessary to create temporary imputed values where the original lifetime usage indicator was missing. The lifetime indicator for smokeless tobacco was used as a covariate for later models, so it was necessary to create this provisional value. In the first provisional hot-deck step, matching was done on a single predicted mean from the PRD step, and missing values were replaced with the value from a donor within this neighborhood.

If possible, donors and recipients were required to be from states with the same level of smokeless tobacco usage (state rank¹¹⁷), where the level of usage was defined in terms of the

¹¹⁵ See Section 3.4.2 for a brief discussion of how order is determined for imputation sets that use the multiple RP/multiple PRD or single RP/multiple PRD type of PMN.

¹¹⁶ SAS®-callable SUDAAN® was used to fit all dichotomous and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in RTI International (2013). SAS software is a registered trademark of SAS Institute Inc. SUDAAN is a registered trademark of Research Triangle Institute.

¹¹⁷ See Section 3.8 for a general discussion of state-rank variables.

weighted proportion of a given state's residents who were lifetime users of the substance.¹¹⁸ If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) the state-rank constraint was removed, and the delta constraint was reapplied.

No logical constraints were placed on the neighborhoods for any of the lifetime usage indicators.

6.5.1.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Substances

PRD models and provisional hot-deck steps were completed in a manner similar to that described above for cigars, pipes, alcohol, marijuana, cocaine and crack, and heroin, with the following deviations:

- Because cocaine and crack were in two separate back-to-back sections in the 2023 NSDUH questionnaire, separate models were fit for the two substances. However, crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood with multivariate matching.¹¹⁹ This was true regardless of whether the item nonrespondent was missing only crack, only cocaine, or both crack and cocaine. Once the neighborhood was defined, missing values for crack and/or cocaine were replaced with the values from one donor within this neighborhood.
- For respondents missing cocaine or crack but not both cocaine and crack, donors had to match respondents on the nonmissing value. If the respondent was a lifetime user of cocaine but missing crack, the donor also had to be a lifetime user of cocaine. If the respondent was a lifetime nonuser of crack, then the donor also had to be a lifetime nonuser of crack. It was not possible for a respondent to be a lifetime nonuser of cocaine and have a missing value for lifetime use of crack, because crack is a type of cocaine. Similarly, it was not possible for a respondent to be a lifetime user of crack and have a missing value for lifetime use of cocaine. This constraint was applied during the first two attempts to find a donor but was dropped in subsequent attempts.
- Heroin did not undergo a provisional imputation step, because it was the last variable in the imputation set.

6.5.1.1.5 Final Hot-Deck Step

Appendix E provides details on the final hot-deck step for the lifetime use indicators for imputation set 1. Although the predictive mean vector could be large if several indicators were missing, the hot-deck step included no logical constraints.

¹¹⁸ Those with a missing lifetime use indicator for the drug were treated as lifetime nonusers.

¹¹⁹ This provisional hot-deck program actually uses a multivariate delta constraint, though the distance from donor to recipient is based only on the predicted probability of lifetime use of cocaine. This was done to avoid the complexity of the Mahalanobis distance calculation. Specifically, this is neither univariate nor multivariate matching.

No final imputation-revised variables indicating lifetime usage alone were created, because this information was recorded in the final imputation-revised recency-of-use variables. Imputation indicators also were not created, though temporary variables to indicate that lifetime usage was imputed were maintained to inform the creation of the recency-of-use imputation indicators.

6.5.1.2 Hallucinogen Child Drug Lifetime Use (Imputation Set 22)

Imputation for the hallucinogen child lifetime use variables for imputation set 22 was done in a manner similar to that described above for set 1. The following deviations from the process described for set 1 applied to set 22.

- **Imputation Order:** The seven hallucinogen child drugs were imputed in the following order: LSD, PCP, Ecstasy, ketamine, the tryptamine hallucinogens DMT/AMT/"Foxy," *Salvia divinorum*, and "other" hallucinogens (includes hallucinogens other than those listed).
- **Response Propensity Step:** The eligible population for the models included lifetime hallucinogen users based on the imputation-revised lifetime variable (imputation set 23). For individuals to be considered item respondents, they must have complete data for lifetime use of LSD, PCP, Ecstasy, ketamine, the tryptamine hallucinogens DMT/AMT/"Foxy," *Salvia divinorum*, and "other" hallucinogens.
- **First Prediction Step:** This step was identical to the process described for set 1.
- **First Provisional CSI Step:** Instead of undergoing a provisional hot-deck step to determine provisional values for use in covariates for later models, the lifetime hallucinogen child drugs underwent a provisional CSI step. See Section 3.6.3 for a general description of the CSI step. Briefly, the PRD model is followed by an imputation step where the predicted probability of lifetime use from the PRD step is used to stochastically impute a value of "Yes" or "No" for each item nonrespondent.
- **Analogous Prediction and Provisional CSI Steps for Remaining Drugs:** PRD models and provisional CSI steps were completed in the manner described above for PCP, Ecstasy, ketamine, the tryptamine hallucinogens DMT/AMT/"Foxy," *Salvia divinorum*, and "other" hallucinogens with one deviation: "Other" hallucinogens did not undergo a provisional imputation step, because it was the last variable in the imputation set.
- **Final Hot-Deck Step:** Appendix E provides details on the final hot-deck step. A single logical constraint was applied requiring donors to be lifetime users of at least one hallucinogen child drug that the recipient is missing if the recipient is a nonuser of all nonmissing hallucinogen child drugs or is missing all child drugs. As with set 1, no final imputation-revised variables indicating lifetime usage alone were created.

6.5.1.3 Imputation-Revised Cigarette Recency and Frequency of Use (Imputation Set 2)

As indicated in [Table 6.7](#), the second set of substance use variables to undergo imputation were the cigarette recency and 30-day frequency variables. The multiple RP/multiple PRD type of PMN was used to process these variables.

6.5.1.3.1 *Sequence of Imputation*

Because recency-of-use and frequency-of-use variables for a given substance were in the same imputation set, the calculation of predicted means for the frequency-of-use variables required the item nonrespondents to be identified as provisional past month and/or past year users. For this reason, cigarette recency was modeled prior to 30-day frequency, and provisional imputations were performed to allow for the identification of provisional past month users of cigarettes.

6.5.1.3.2 *First Response Propensity Step (Cigarette Recency of Use)*

The input to the weight adjustment model in the RP step for imputation set 2 was the preliminary analysis weight. To impute for missing recency-of-use values for cigarettes, it was first necessary to define the domain within each of the three age groups. Using the imputation-revised lifetime indication of use, the file was reduced to lifetime cigarette users.¹²⁰ Among these lifetime users, item respondents and nonrespondents were identified across recency-of-use and 30-day frequency-of-use measures. If a valid response was provided for each substance use measure, then individuals were deemed item respondents for cigarettes. Otherwise, they were item nonrespondents. See Appendix D for information on the covariates used in the RP models for this variable.

6.5.1.3.3 *First Prediction Step (Cigarette Recency of Use)*

Using the nonresponse adjusted weights, the probability of selecting each cigarette recency-of-use category was modeled within each age group using polytomous logistic regression. SUDAAN's MULTILOG procedure was used to estimate the parameters from the appropriate logistic model from which predicted probabilities for each of the cigarette recency categories were calculated for both item respondents and item nonrespondents. The four recency categories were the following:

1. past month;
2. past year, not past month;
3. past 3 years, not past year; and
4. lifetime, not past 3 years.

6.5.1.3.4 *First Provisional Hot-Deck Step (Cigarette Recency of Use)*

To define the domain for the cigarette 30-day frequency-of-use variable, it was necessary to create temporary imputed values where the original cigarette recency value was missing. In order to save time and conserve resources, and because the imputation was only provisional, a univariate matching procedure was implemented. The only predicted mean used was the predicted probability of past month use, because past month use was the most critical measure of recency of cigarette use.

¹²⁰ No imputation was required for lifetime use of cigarettes due to the usable interview rule. See Sections 2.2.3 and 6.5.1.1 for more information on the usable interview rule.

If possible, donors and recipients were required to be from states with the same level of usage of a given substance (state rank; see Section 3.8), where the level of usage was defined in terms of the weighted proportion of a given state's residents who had used cigarettes in the past month.¹²¹ If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) donors and recipients were no longer required to be from states with similar usage levels.

The only logical constraints placed on the neighborhoods involved situations where a general recency category was available for a respondent and imputation was required to determine the specific recency categories. The general recency categories that appeared are discussed in Section 6.2.2.3 and shown in [Table 6.8](#). Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood.

Table 6.8 General Incomplete Recency Categories for Cigarettes

General Incomplete Recency Category	Allowable Specific Recency Categories
Lifetime	1. Past month 2. Past year but not past month 3. Past 3 years but not past year 4. Lifetime but not past 3 years
Past Year	1. Past month 2. Past year but not past month
Lifetime, Not Past Year	1. Past 3 years but not past year 2. Lifetime but not past 3 years
Lifetime, Not Past Month	1. Past year but not past month 2. Past 3 years but not past year 3. Lifetime but not past 3 years
Lifetime, Not Past Month but within Past 3 Years	1. Past year but not past month 2. Past 3 years but not past year
Past 3 Years	1. Past month 2. Past year but not past month 3. Past 3 years but not past year

6.5.1.3.5 Second Response Propensity Step (Cigarette 30-Day Frequency)

The input to the weight adjustment model in the RP step for imputation set 2 was the preliminary analysis weight. The modeling of cigarette 30-day frequency followed that of recency. The file was first reduced to the domain, which was past month cigarette users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the cigarette recency imputations. To be an item respondent, the individual had to have provided valid responses to both the cigarette recency and 30-day frequency measures. The item RP adjustment was then computed so that

¹²¹ Those individuals whose past month use status was unknown were treated as if they were not past month users.

the respondents' weights accurately represented all past month users of cigarettes. See Appendix D for information on the covariates used in the RP models for this variable.

6.5.1.3.6 Second Prediction Step (Cigarette 30-Day Frequency)

As stated in the previous section, only past month users of cigarettes were used to build the 30-day frequency-of-use model. The response variable of interest in the 30-day frequency-of-use models, prior to a normalizing transformation, was the proportion of the days in a month (30 days) on which a respondent used cigarettes. The range of values for the proportion was from (greater than) 0 to 1. Hence, to model 30-day frequency of use, the following empirical logit transformation was computed for all respondents:

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

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

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

where P_i was defined as the proportion of days in the past month on which respondent i used the substance. The standard logit transformation was not used because it was not defined for daily users.¹²² Using the adjusted weights, a linear univariate regression model was then fitted using SUDAAN software for the log-transformed variable Y_i within each age group.

For cigarettes, the empirical distribution for 30-day frequency of use was in fact a mixture distribution, with a positively skewed distribution from 1 to 29 and a spike at 30. This substance was modeled using two separate models. One was a logistic model for daily use versus nondaily use among past month users. For the nondaily past month users (i.e., those who had used between 1 and 29 days), the model described above was used. In this instance, the response variable in a linear regression model was a logit of the proportion of the period (30 days) during which a respondent used the substance. Both the predicted probability of daily use and the logit of the proportion of the month used (assuming nondaily use) were used as predicted means in the subsequent hot-deck step. The logit was back-transformed into a proportion before use in the hot-deck step.

6.5.1.3.7 Final Hot-Deck Step (Cigarette Recency and 30-Day Frequency)

The full predictive mean vector for cigarettes contained probabilities associated with several of the recency-of-use categories, a probability of daily use, and a predicted probability of use on a given day in the past month. Each element in the full vector of predicted means was adjusted so that all elements were conditioned on the same usage status whenever possible. The

¹²² See Cox and Snell (1989) for a discussion of the empirical logit transformation. Here, if the respondent was a daily user of the substance, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$, with $N = 30$ so that it was defined for all respondents.

elements in the predictive mean vector that could have potentially resulted are shown in [Table 6.9](#), with the assumption that only the lifetime usage is known. If other information about the recency of use is known (e.g., past year user), the predictive mean vector is adjusted accordingly. The portion of the full predictive mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If partial information was available regarding recency of use, then that information was used to adjust the recency-of-use probabilities. The portions of the full predictive mean vector used for each missingness pattern, with accompanying adjustments, are provided in Appendix E. The Mahalanobis distance was then calculated using only the portion of the predictive mean vector that was associated with the given missingness pattern, with elements appropriately adjusted. The likeness and logical constraints applied to each missingness pattern are also available in Appendix E.

Table 6.9 Elements of Full Predictive Mean Vector: Cigarettes

Substance Use Measure and Category of Interest	Predicted Mean
Recency of Use, Past Month Use ¹	$P(\text{past month user} \mid \text{lifetime user})$
Recency of Use, Past Year but Not Past Month Use ¹	$P(\text{past year but not past month user} \mid \text{lifetime user})$
Recency of Use, Past 3 Years but Not Past Year Use ¹	$P(\text{past 3 years but not past year user} \mid \text{lifetime user})$
30-Day Frequency of Use for Nondaily Users over Past 30 Days	$P(\text{use on a given day in the month} \mid \text{past month user, not a daily user}) \times P(\text{not a daily user} \mid \text{lifetime user}) \times P(\text{past month user} \mid \text{lifetime user})^2$
Daily User over Past 30 Days	$P(\text{daily user} \mid \text{past month user}) \times P(\text{past month user} \mid \text{lifetime user})^2$

¹ The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.

² Interpreting the proportion of the month used as a probability of use on a given day in the month requires analysts to assume that the probability of use on each day in the month was equal, which is not true.

6.5.1.3.8 Final Variables (Cigarette Recency and 30-Day Frequency)

The final imputation-revised recency-of-use and 30-day frequency variables were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the substance (CIG) and the final two letters identified the measure (RC = recency of use, FM = 30-day frequency). Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR. The levels for the imputation indicators were the standard levels used for most imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user).

6.5.1.4 Imputation-Revised Cigarette Age at First Use (Imputation Set 3)

As indicated in [Table 6.7](#), the third imputation set consisted of the cigarette AFU variable. Recent initiates of substance use (i.e., those whose current age was equal to or 1 year greater than the reported AFU) were asked the year and month of their first use. To have this

information for all users, both missing MFU and missing YFU for less recent initiates (and recent initiates who did not report MFU and YFU) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised AFU, and imputation-revised recency and frequency variables. To have complete date-of-first-use information, day of first use (DFU) was randomly assigned for all users. The combined data gave the respondent's AFU along with the date of first use.

6.5.1.4.1 Response Propensity Step (Cigarette Age at First Use)

The input to the weight adjustment model in the RP step for imputation set 3 was the preliminary analysis weight. To impute for missing cigarette AFU, it was necessary to define the eligible population. Using the imputed recency of use, the files were reduced to lifetime users of cigarettes. If a valid response was provided for the AFU measure,¹²³ the individual was deemed an item respondent. See Appendix D for information on the covariates used in the RP models for this variable.

6.5.1.4.2 Prediction Step (Cigarette Age at First Use)

The response variable in the prediction model for AFU, before a normalizing transformation, was the AFU as a proportion of the current age. The numerator in this proportion was an integer representing AFU. However, because this integer was in fact a truncated version of the real AFU, the value was made continuous by adding a random component between 0 and 1. Hence, expressing the proportion as $P_i = Y_i/N_i$, the numerator was given as

$$Y_i = \text{Age at First Use}_i + \text{Uniform}(0,1) \text{ random number}.$$

In the event that the age at first use was equal to the age, Y_i was constrained so that it was equally likely to be anywhere on the interval $[\text{Age at First Use}_i, N_i]$. Thus, Y_i was prevented from being greater than N_i . The denominator in the proportion was the total age. The true age was known, based on the interview date and birth date. Expressing it in years rather than days required dividing by the number of days in the year:

$$N_i = (\text{Interview Date} - \text{Birth Date} + 1)/365.25.$$

After a weight adjustment, the empirical logit transformation was used as the response variable in a weighted linear univariate regression:

$$\log\left[\frac{(Y_i + 0.5)}{(N_i - Y_i + 0.5)}\right].$$

This transformation was nearly equivalent to the standard logit transformation:

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

¹²³ Respondents who reported an age at first use of 1 or 2 were treated as item nonrespondents in the RP and PRD steps, because of the implausibility of such a young age at first use. In the hot-deck step, their response was left unchanged, but they were not allowed to be donors.

which was not used, because it might be unstable for respondents who started using at their current age.

One unusual covariate used in the PRD model for cigarette AFU was a modified 30-day frequency variable for cigarettes. It was defined as follows:

$$\begin{aligned} \text{new30}_i &= 0 && \text{if respondent } i \text{ did not use cigarettes in the past month and} \\ &= 30\text{-day frequency} && \text{if respondent } i \text{ used cigarettes in the past month.} \end{aligned}$$

Naturally, the full model for AFU did not include the lifetime indicator for the substance in question, because the model was built on cigarette users. Information on the covariates used in these models can be found in Appendix D.

From the final model, a predicted value (based on the Y variable) was calculated for each cigarette user, which was then back-transformed to produce a predicted cigarette AFU.

6.5.1.4.3 *Hot-Deck Step*

The imputation-revised cigarette AFU assignment was conducted using a single predicted mean: the predicted AFU. Appendix E provides a complete list of likeness and logical constraints applied to the cigarette AFU imputations. The likeness constraints for AFU were more stringent than those for the other substance use measures. Therefore, it was often necessary to loosen the constraints. Once these likeness constraints were removed, some complex logical constraints remained, based on the interview date, the birth date, and imputation-revised recency and frequency values.

6.5.1.4.4 *Date-of-First-Use Assignments*

After the AFU imputations, all lifetime users of cigarettes had nonmissing AFU values. Using this AFU, users were assigned values for MFU, YFU, and DFU. Recent initiates, or those respondents whose AFU was within 1 year of their age, were asked for their MFU and YFU. The DFU was not collected in the questionnaire and was missing for all respondents. The MFU, YFU, and DFU data contained four patterns of missingness:

Pattern 1: Recent initiates: missing day of first use only;

Pattern 2: Recent initiates: missing month/day of first use;

Pattern 3: Recent initiates: missing year/month/day of first use; and

Pattern 4: Less recent initiates: missing year/month/day of first use.

For each missingness pattern, upper and lower bounds on the date of first use (i.e., the earliest possible date of first use and the latest possible date of first use) were determined. Once the earliest and latest possible dates of first use were determined, a day was randomly selected from this interval. The imputation-revised month/day/year values were then extracted from this date of first use.

6.5.1.4.4.1 Missingness Pattern 1

In this missingness pattern, a recent initiate provided all the information asked by the questionnaire (i.e., both the MFU and YFU). However, to obtain a complete date of first use, a DFU also was needed. Thus, a DFU was randomly assigned, given the respondent's MFU and YFU, in a way that was consistent with both the 30-day frequency/recency and AFU. Below is a brief description of the process used to obtain a date of first use in such situations. The imputed YFU, MFU, and DFU were extracted from the date, as defined below:

Final date of first use = Earliest possible date + [(Days between earliest and latest date) × (a random number generated from a Uniform (0,1) distribution)],

where

Days between earliest and latest date = Latest possible date – Earliest possible date + 1;

Earliest possible date = maximum [(AFUth birthday), (first day of the month indicated by MFU/YFU)]; and

Latest possible date =

- minimum [(Interview date – 30-day frequency + 1), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)], if recency = 1;
- minimum [(Interview date – 30), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)], if recency = 2; and
- minimum [(Interview date – 1 year), (1 day before the (AFU + 1)th birthday), (last day of the month indicated by MFU/YFU)], if recency = 3.

Note that it is impossible for recent initiates to have recency = 4 (lifetime but not past 3 years). Recent initiates had to have begun using the substance no earlier than their (AFU)th birthday. Because AFU = current age, or AFU = current age – 1, their (AFU)th birthday was within the past 2 years. Respondents who had begun using the substance within the past 2 years must logically have last used the substance within the past 2 years, and therefore could not have had recency = 4.

Occasionally, the earliest possible date was set to 29 days before the interview. This occurred for respondents meeting all of the following conditions:

1. The latest possible date was within 29 days of the interview.
2. The earliest possible date determined by the above rule was within a year of the interview.
3. The recency = 1.
4. The 12-month frequency = 30-day frequency (if applicable), or the 12-month frequency = 1.

Logically, all the lifetime usage of the substance for these respondents occurred in the past 30 days (including the interview date). The first condition ensures that the application of this rule does not cause an inconsistency. The second condition implies that the substance was not used by these respondents more than 1 year ago. The third and fourth conditions imply that the substance was not used by these respondents in the interval (1 year before the interview, 1 month before the interview). Therefore, these respondents did not use the substance more than 1 month ago. All their lifetime use must have occurred in the past month.

6.5.1.4.4.2 Missingness Pattern 2

The second missingness pattern occurred when recent initiates provided their YFU but did not provide an MFU. In such instances, a month and day were randomly assigned that were consistent with both the respondent's frequency/recency and with the AFU range. The imputed MFU and DFU were derived in the same manner as the date of first use in Missingness Pattern 1, except with the following changes:

- For the earliest possible date, replace "first day of the month indicated by MFU/YFU" with "January 1st of the YFU."
- For the latest possible date, replace "last day of the month indicated by MFU/YFU" with "December 31st of the YFU."

6.5.1.4.4.3 Missingness Pattern 3

Similar to Missingness Pattern 2, the third missingness pattern occurred when recent initiates provided neither an MFU nor a YFU value. The MFU, YFU, and DFU were randomly assigned from a uniform distribution in a way that was consistent with both the cigarette 30-day frequency/recency and the AFU. Again, the imputed MFU, YFU, and DFU were derived in the same manner as described in Missingness Pattern 1.

6.5.1.4.4.4 Missingness Pattern 4

The fourth missingness pattern occurred when respondents reported, or were imputed to, an AFU at least 2 years less than their age. In this missingness pattern, the 30-day frequency was immaterial to the final date of first use because the respondent could not have begun using in the past month:

Earliest possible date = AFUth birthday; and

Latest possible date =

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

6.5.1.4.5 *Final Age and Date-of-First-Use Variables*

The final imputation-revised date-of-first-use variables were identified with the prefix IR, followed by a six-letter identifier, where a three-letter code identified the substance¹²⁴ and the final three letters identified the measure (AGE = age at first use, MFU = month of first use, YFU = year of first use, DFU = day of first use). Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were the standard levels used for most imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user). Because survey respondents are not asked for the specific day on which they first used the substance of interest, all respondents in the domain receive IIXxxDFU = 3. Also, as indicated above, only recent initiates are asked for the year and month of first substance use. Subsequently, these questions have high rates of nonresponse because of the skip logic embedded in the questionnaire, as all other individuals in the domain require imputation for their MFU and YFU.

6.5.1.5 **Imputation-Revised Age at First Daily Cigarette Use (Imputation Sets 4 and 5)**

In addition to AFU, the cigarettes section also included a question asking for the respondent's age at first daily cigarette use, where a daily user was defined as someone who reported having at some time smoked cigarettes every day for a period of at least 30 days. Imputation procedures for age at first daily cigarette use were similar to AFU, with some key exceptions as discussed below.

One such exception involved the domain of the AFU variable. Whereas the AFU question was asked of all cigarette users, the age-at-first-daily-use question was asked of only daily users. The "daily use" indication came from two sources. If respondents answered either the 30-day frequency or estimated 30-day frequency with a "30," or if they had a "yes" value for the edited variable associated with the "ever daily used" question (CIGDLYMO), then they were considered to be daily users. For more information about CIGDLYMO, see Section 6.2.6.

Anyone who answers the 30-day frequency or estimated 30-day frequency with a "30" is automatically skipped out of this question. Therefore, it is necessary to obtain the imputation-revised cigarette 30-day frequency (IRCIGFM) prior to imputing the lifetime-daily-cigarette-use variable (IRCDULF) so that it is not included with the other lifetime substance use indicators as part of imputation set 1. Instead, as indicated in [Table 6.7](#), the age at first daily cigarette use actually contains two separate imputation sets. Imputation set 4 includes the lifetime indicator of daily cigarette use and imputation set 5 includes the age-at-first-daily-cigarette-use variables. At this stage in the process, there should be no missing responses to the 30-day frequency question, which were imputed as part of imputation set 2 as discussed above. Daily users,

¹²⁴ Exceptions to this rule occurred with marijuana and cigarette daily use. For historical reasons, marijuana contained a two-letter code (MJ). Marijuana variables therefore ended with a five-letter identifier rather than a six-letter identifier. The code for cigarette daily use was CDU, which differed from the general cigarette code of CIG. Details about cigarette daily use are provided in Section 6.5.1.5.

based on 30-day frequency, should be either known (based on a response in the survey) or imputed. However, responses for the ever-daily-used question (CIGDLYMO) could still be missing, and, therefore, it was first necessary to impute these values to define the domain for the age-at-first-daily-use variable.

6.5.1.5.1 Response Propensity Step (Ever-Daily-Used Cigarettes)

The input to the weight adjustment model in the RP step for imputation sets 4 and 5 was the preliminary analysis weight. To impute for missing values in the ever-daily-used variable, it was necessary to define the domain: lifetime users of cigarettes who had an imputation-revised 30-day frequency¹²⁵ of fewer than 30 days (includes legitimate skip codes for lifetime but not past month users). If a valid response was provided in the ever-daily-used variable, the individual was deemed an item respondent. See Appendix D for information on the covariates used in the RP models for this variable.

6.5.1.5.2 Prediction Step (Ever-Daily-Used Cigarettes)

After the weights were adjusted, the ever-daily-used variable was modeled using weighted logistic regression in SUDAAN. The predicted mean from this model was the predicted probability of ever smoking cigarettes daily.

6.5.1.5.3 Hot-Deck Step (Ever-Daily-Used Cigarettes)

The predicted mean from the prior step was used in a straightforward hot-deck step, which is summarized in Appendix E.

6.5.1.5.4 Hot-Deck Step (Age at First Daily Cigarette Use)

Instead of separately modeling age at first daily cigarette use, the predicted means from the age-at-first-cigarette-use models were used to determine neighborhoods. The imputation-revised age-at-first-daily-use assignment was conducted using univariate matching and univariate assignment.

All the logical constraints applied to age at first cigarette use were also applied to age at first daily cigarette use. Besides those logical constraints, an additional logical constraint was applied specifically to age at first daily cigarette use. If the cigarette AFU was not missing for a recipient with a missing age at first daily use, the donors were prevented from having an age at first daily use earlier than the preexisting cigarette AFU. This constraint was applied as daily cigarette users constitute a subset of all cigarette users. Therefore, daily use of cigarettes can be thought of as a child measure for cigarettes, with a lifetime indicator and an AFU measure but no recency or frequency. This association required that these variables remain internally consistent for each respondent.

The likeness constraints were nearly identical to those used for cigarette AFU. There was only one difference: an additional step was employed if no donor was found after loosening all of the

¹²⁵ The imputation-revised 30-day frequency included responses from the 30-day frequency question (CG07), as well as the estimated 30-day frequency question (CG07DKRE).

likeness constraints. In particular, if the AFU and age at first daily use were both initially missing, the imputed AFU was set back to missing and reimputed simultaneously with the age at first daily use so that they were mutually consistent.¹²⁶

Subject to these constraints, the age-at-first-daily-use value of the randomly selected donor was then assigned to the recipient.

6.5.1.5.5 *Assignment of Date of First Daily Cigarette Use*

After the imputation-revised age at first daily cigarette use was created, all daily cigarette users had a valid age at first daily cigarette use. From this age, a year/month/day of first daily use was assigned. The date assignment procedure was identical to the procedure described in Section 6.5.1.4.4 with the following exception. In the setting of the earliest possible date for daily cigarette use, the overall cigarette date of first use was used as an additional bound. This was done for cigarettes and other substances to ensure that the child date of first use was never earlier than the parent date of first use.

6.5.1.5.6 *Final Variables*

The final imputation-revised date-of-first-daily-cigarette-use variables were named in the same manner as described in Section 6.5.1.4.5. However, the three-letter identifier for cigarette daily use was CD2. As with the general cigarette use variables, each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

6.5.1.6 *Recency and Frequency of “Other” Substances*

Imputation for the parent and child recency and frequency variables for imputation sets 6, 8, 10, 11, 13, 17, and 19 in [Table 6.7](#) (smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine and crack, and heroin, respectively) was done in a manner similar to that described above for cigarettes. The following deviations from the process described for cigarettes applied to these “other” substances.

The order of imputation for smokeless tobacco and cigars was identical to that for cigarettes, with recency of use being modeled first, followed by 30-day frequency. However, not all imputation sets included the same variables. Alcohol, marijuana, cocaine and crack, and heroin also included a measure for 12-month frequency of use. For these substances, imputation proceeded in the following order: recency of use, 12-month frequency of use, and 30-day frequency of use. For a given substance, this ordering allowed recency of use to be included in the model for 12-month frequency of use and allowed 12-month frequency of use to be included in the model for 30-day frequency. Further, this ordering allowed the provisional recency of use to define the domains for the frequency measures. For pipes, the respondents were only asked about their most recent use, and no information was collected regarding

¹²⁶ Though it has occurred in prior years, the situation where no donors were available, even after loosening all constraints, did not occur in the 2023 NSDUH.

frequency of use in the past year or month. Therefore, only the recency-of-use variable required modeling and imputation.

6.5.1.6.1 Recency of Use

6.5.1.6.1.1 Response Propensity Step

The input to the weight adjustment model in the RP step for the parent and child recency-of-use variables was the preliminary analysis weight. Similar to cigarettes, the eligible population for the recency-of-use models included all lifetime users of the substance of interest as identified by the imputation-revised lifetime substance use variables. However, the identification of respondents and nonrespondents for the purpose of imputation differed from substance to substance depending on the information collected in the questionnaire. In general, an individual had to have provided a valid response to all variables included in the imputation set to be classified as a respondent. If the imputation set included both “parent” and “child” measures ([Table 6.1](#)), then this requirement extended across all observed measures of substance use. For example, to be classified as a respondent for the cocaine and crack imputation set, a valid response must be provided for recency of use, 12-month frequency of use, and 30-day frequency of use for both the cocaine and crack questions. See Appendix D for information on the covariates used in the RP models for these variables.

6.5.1.6.1.2 Prediction Step

Only cigarettes, cigars, and smokeless tobacco included a recency category for past 3 years but not past year. For all other substances except pipes, the outcome variable was a three-level categorical variable with the following levels:

- past month;
- past year, not past month; and
- lifetime, not past year.

For pipes, the outcome variable had only two levels:

- past month; and
- lifetime, not past month.

Because cigarettes were the first recency/frequency imputation set, it was not possible to include the recency information for other substances as covariates in the PRD model. However, for substances other than cigarettes, recency-of-use covariates for cigarettes, alcohol, and marijuana replaced the lifetime indicators where applicable. For example, the PRD model for alcohol included recency indicators for cigarettes but only included the lifetime usage indicator for marijuana because this substance comes later in the sequence.

6.5.1.6.1.3 Provisional Hot-Deck Step

Sometimes a general recency category was assigned during the editing process, and the specific recency was then determined during imputation. However, the categories available for both the general recency and the specific recency varied from substance to substance

depending on the number of levels included in the recency-of-use measure. The allowable general and specific recency categories for cigarettes are shown in [Table 6.8](#). The same categories apply to cigars and smokeless tobacco. For all other substances except marijuana and pipes, the only general incomplete recency categories that were applicable were lifetime and past year (the first two rows). For marijuana, the only general incomplete recency categories that were applicable were lifetime, past year, and lifetime, not past month (the first two rows and the fourth row). For pipes, only the lifetime category was applicable. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood of potential donors.

Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. No provisional imputed values were determined for substances that were a subset of the substance associated with the predicted mean (parent/child measures). This occurred with cocaine (parent) and crack (child).

6.5.1.6.1.4 Hierarchical Modeling

For heroin, the proportion of users who were past year users was quite small when compared with the total number of lifetime users. The lopsided distribution^{[127](#)} for heroin caused convergence problems when fitting polytomous logistic models. To alleviate this problem, the single polytomous logistic model was replaced with two dichotomous logistic models^{[128](#)} that were fit hierarchically.

As with the polytomous logistic model, the first dichotomous logistic model was fit among lifetime users, but the past month and past year but not past month categories in the response variable were collapsed into a single level. In a similar manner to other recency-of-use models, respondents' weights were adjusted so that they represented all lifetime users. The predicted probability of past year use given lifetime use was obtained from this model.

The second model was limited to past year users, where the response variable had two levels: past month and past year but not past month users. For the second model, respondents' weights were adjusted so that they represented all past year users. In order to do this, it was necessary to completely define the domain of past year users. Missing values were provisionally imputed to past year or not past year use by randomly allocating the response using the predicted means from the first model.

From the two dichotomous logistic models, both the probability of past month use and the probability of past year but not past month use were obtained and used in the provisional hot-deck program for recency. Once the predicted means were determined from the two models,

^{[127](#)} A "lopsided distribution" in the context of recency of use is where, among the categories of past month use, past year but not past month use, and lifetime not past year use, only a small minority of respondents gave a response of "past month use."

^{[128](#)} The set of covariates used for these dichotomous logistic models were the same as those for logistic modeling given earlier in this section.

a single vector of predicted means conditional on lifetime usage, as with the polytomous logistic models, was determined as follows:

$$P(\text{past month use} \mid \text{lifetime use}) = P(\text{past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}), \text{ and}$$

$$P(\text{past year, not past month use} \mid \text{lifetime use}) = P(\text{past year, not past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}).$$

6.5.1.6.2 12-Month Frequency of Use

The modeling of 12-month frequency sequentially followed that of recency of use for alcohol, marijuana, cocaine and crack, and heroin.

6.5.1.6.2.1 Response Propensity Step

The input to the weight adjustment model in the RP step for the 12-month frequency-of-use variables was the preliminary analysis weight. The eligible population for the imputation of 12-month frequency of use was past year users of the substance in question (as defined by the provisional recency of use). The item response indicator and the RP adjustment were defined among the past year users of each substance. Item respondents were defined using the same criterion as was used in the recency-of-use imputations. Namely, the respondent had to have a valid response to all of the applicable measures for the substance of interest. The item RP adjustment was then computed so that the respondents' weights accurately represented all past year users of the substance. See Appendix D for information on the covariates used in the RP models for these variables.

6.5.1.6.2.2 Prediction Step

As indicated in the previous section, only past year users of the substance of interest were used to build the 12-month frequency-of-use model. The response variable of interest in the 12-month frequency-of-use models for most respondents, prior to a normalizing transformation, was the proportion of the days in a full year (365.25) on which a respondent used a particular substance. For example, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/365.25. Some respondents, however, started using the substance within the past year. If they responded to the MFU question, the difference between the MFU and the date of the interview indicated the total time period during which they could have been using the substance of interest.¹²⁹ If the date of the interview was July 10, for example, and the MFU was March of the same year, the maximum period during which the respondent could have used is the number of days between March 1 and July 10 (inclusive), or 101. Thus, if a respondent entered a 12-month frequency of 100, the

¹²⁹ If respondents initiated use in the past year (according to their age-at-first-use or year-of-first-use response), but they did not answer the month-at-first-use question, then the maximum period the respondents could have been using the substance was the number of days between the interview date and the date of the respondents' last birthdays or January 1 of the survey year (if respondents reported first using in the survey year), whichever yielded the smallest number of days.

(untransformed) response variable of interest would be 100/101 instead of 100/365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\frac{(Y_i + 0.5)}{(N_i - Y_i + 0.5)}\right],$$

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

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

where P_i is defined as the proportion of days in the past year in which respondent i used the substance. The standard logit transformation was not used because it was not defined for daily users.¹³⁰ Using the adjusted weights, a linear univariate regression model using SUDAAN software was then fitted for the log-transformed variable Y_i within each age group.

- **Covariates:** Because the 12-month frequency models were limited to past year users, only two recency categories could have resulted: past month use and past year but not past month use.¹³¹ Hence, recency of use for the substance being modeled was represented as a covariate in the 12-month frequency-of-use model by a single indicator variable representing these two categories. Imputation-revised recency of use for other substances was used if available. If the missing values for a given substance's recency of use had not yet been imputed, a single covariate was used that indicated lifetime usage of that substance. To control for state variations in substance use, the state-rank groups defined for the recency-of-use imputations were included as covariates in the 12-month frequency-of-use models. Appendix D provides information on the covariates used in the 12-month frequency-of-use models.
- **Predicted Means:** The predicted mean that resulted from the 12-month frequency-of-use model was a logit of the proportion of the year used. This logit was back-transformed into a proportion for use as the variable from which the neighborhoods were created. This proportion could be treated as a probability, which, in turn, could be multiplied by the probability of past year use to make the predicted mean conditional on lifetime use of the substance in question. When calculating predicted means for some item nonrespondents, sometimes it was not known whether they were past year users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

¹³⁰ See Cox and Snell (1989) for a discussion of the empirical logit transformation. If the respondent was a daily user of the substance, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$, with $N = 365.25$ so that it was defined for all respondents.

¹³¹ For item nonrespondents, where parameter estimates were used to determine predicted means, past year use was defined based on a provisional imputation.

6.5.1.6.2.3 Provisional Hot-Deck Step

For imputation sets that included both 12-month frequency and 30-day frequency—alcohol (Set 11), marijuana (13), cocaine and crack (17), and heroin (19)—it was necessary to provisionally impute the 12-month frequency-of-use variable so that it could be used as a covariate in the 30-day frequency-of-use imputations.

The logical constraints involved the interview date, initiation data (AFU, MFU, and YFU), birthday, recency of use, and 30-day frequency of use. The likeness constraints used in the assignment of values for 12-month frequency of use were similar to those used for recency of use. State-rank groups were again based on level of past month usage. Recipients and donors were also required to have the same recency of use (past month vs. past year but not past month), whether that recency of use was reported or imputed.¹³² If no donors were available within these constraints, then they were loosened in the following order: (1) the delta constraint was removed, (2) donors and recipients were no longer required to be from states with similar usage levels, and (3) donors and recipients were no longer required to have the same recency of use.

6.5.1.6.2.4 Assignment of Provisional Imputed Values

For all substance use measures except 12-month frequency, the observed value of interest was donated directly to the recipient. However, because donors and recipients could have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of the total period was donated rather than the observed 12-month frequency. In the assignment step, the donor's proportion of the total period was multiplied by the corresponding recipient's maximum possible number of days in the year on which the recipient could have used the substance in order to arrive at a 12-month frequency-of-use value.

Occasionally, more than one substance was associated with a single predicted mean. However, for the provisionally imputed values, only the parent measure was used as a covariate in later models. Therefore, multivariate assignments were not needed in the provisional hot-deck step, but they did occur in the final hot-deck step for recency and frequency. For example, the recency and frequency variables for cocaine and crack formed a single imputation set (17). Although 12-month frequency questions were asked for both cocaine and crack, only the 12-month frequency for cocaine was modeled, and only the 12-month frequency for cocaine was used as a covariate in the subsequent PRD model (30-day cocaine frequency). This means that there was no need to impute a provisional value for 12-month frequency for crack.

¹³² Because all respondents in the 12-month frequency-of-use imputation were past year users by definition, item nonrespondents who were past month users required donors who were past month users, and item nonrespondents who were past year but not past month users required donors who matched that specific recency category.

6.5.1.6.3 30-Day Frequency of Use

6.5.1.6.3.1 Response Propensity Step

The input to the weight adjustment model in the RP step for the 30-day frequency-of-use variables was the preliminary analysis weight. As with cigarettes, the file was first reduced to the domain of past month users, as defined by the provisionally imputed recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the recency and 12-month frequency imputations. To be an item respondent, the individual had to have provided valid responses to all applicable measures for the substance of interest. The item RP adjustment was then computed so that the respondents' weights accurately represented all past month users of the substance. In contrast with the RP model for cigarettes, the provisional 12-month frequency was included as a covariate for those substance use sections that asked the respondent to report this measure ([Table 6.7](#)). In addition, recencies of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, and heroin were included if available.¹³³ See Appendix D for information on the covariates used in the RP model for these variables.

6.5.1.6.3.2 Prediction Step

As described for cigarettes, only past month users of the substance of interest were used to build the 30-day frequency-of-use model. The response variable of interest in the 30-day frequency-of-use models, prior to a normalizing transformation, was the proportion of the days in a month (30 days) on which a respondent used the substance of interest. In contrast with cigarettes, all other substances that included a measure for 30-day frequency used a single model for all past month users rather than two separate models for daily use and nondaily use.

6.5.1.6.3.3 Provisional Hot-Deck Step

Because there were no substances within the multivariate set of recency-of-use and frequency-of-use measures lower in the sequence than 30-day frequency, there was no need to provisionally impute 30-day frequencies.

6.5.1.6.4 Final Hot-Deck Step

The same principles that applied to the final hot-deck step for cigarettes applied to other substances. However, for substances with child measures and substances with both 12-month and 30-day frequencies, the logical constraints were considerably more complicated, the predictive mean vectors were larger, and the number of missingness patterns was greater. Appendix E provides detailed information on these hot-deck steps, and [Table 6.10](#) provides a listing of the full predictive mean vector as applied to all final hot-deck programs for recency and frequency.

¹³³ If the recency of use for a particular substance was not yet defined, the lifetime indication of use was used instead. The recency of use of the substance being modeled was not used, because all respondents in the model were past month users.

Table 6.10 Elements of Full Predictive Mean Vector

Substance Use Measure and Category of Interest	Predicted Mean	Substance
Recency of Use, Past Month Use ¹	$P(\text{past month user} \mid \text{lifetime user})$	All substances
Recency of Use, Past Year but Not Past Month Use ¹	$P(\text{past year but not past month user} \mid \text{lifetime user})$	All substances except pipes
Recency of Use, Past 3 Years but Not Past Year Use ¹	$P(\text{past 3 years but not past year user} \mid \text{lifetime user})$	Tobacco products ² only
12-Month Frequency of Use	$P(\text{use on a given day in the year} \mid \text{past year user}) \times P(\text{past year user} \mid \text{lifetime user})^3$	All substances except tobacco
30-Day Frequency of Use for Alcohol and Substances with Few Daily Users ⁴	$P(\text{use on a given day in the month} \mid \text{past month user}) \times P(\text{past month user} \mid \text{lifetime user})^5$	All substances except cigarettes and pipes
30-Day Frequency of Use for Substances with Many Daily Users (excluding Alcohol and Smokeless Tobacco)	$P(\text{use on a given day in the month} \mid \text{past month user, not a daily user}) \times P(\text{not a daily user} \mid \text{lifetime user}) \times P(\text{past month user} \mid \text{lifetime user})^5$	Cigarettes
Daily User over Past 30 Days	$P(\text{daily user} \mid \text{past month user}) \times P(\text{past month user} \mid \text{lifetime user})^5$	Cigarettes

¹ The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.

² "Tobacco products" includes cigarettes, cigars, and smokeless tobacco.

³ Interpreting the proportion of the year used as a probability of use on a given day in the year assumed that the probability of use on each day in the year was equal, which was not true. However, the violation of this assumption did not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, but it did allow the predicted mean to be made conditional on what was known.

⁴ Alcohol, with many daily users, was included in this group because the distribution did not show a severe drop-off from 30 days a month to 29 days a month, as was apparent with cigarettes.

⁵ Interpreting the proportion of the month used as a probability of use on a given day in the month assumed that the probability of use on each day in the month was equal, which was not true, in the same manner as the 12-month frequency of use (see note #3 for this table).

The construction of the predictive mean vectors for certain substances was often complex. The main reason for the complexity is that recency and frequency models were not fit for all child measures. In fact, the predicted means from the models for the parent measure were often used as surrogates for the child measure predicted means to reduce the number of models that needed to be fit and to avoid convergence problems with small sample sizes for some of the rarer child measures. For example, if individuals requiring imputation were past year users of cocaine but they had a missing crack recency, then the predictive mean vector includes the probability of past month cocaine use, given that these individuals were past year users of cocaine. When constructing the predictive mean vectors, the following general principles were followed:

1. If both the parent recency and the child recency were missing, condition on the general recency category of the parent measure.
2. Condition all elements of the predictive mean vector on the same general recency level.

6.5.1.6.5 Final Variables

Similar to the final imputation-revised recency-of-use and 30-day frequency variables for cigarettes (IRCIGRC and IRCIGFM), the final imputation-revised recency and frequency variables for other substances with the exception of smokeless tobacco were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the substance and the final two letters identified the measure. In addition to the RC and FM identifiers used for cigarettes, the identifier FY was used for the 12-month frequency variable. For smokeless tobacco, the final imputation-revised recency and frequency variables are IRSMKLSSREC and IRSMKLSS30N. Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.5.1.7 Age at First Use of “Other” Substances

The AFU variables for imputation sets 7, 9, 12, 16, 18, and 20 in [Table 6.7](#) (smokeless tobacco, cigars, alcohol, marijuana, cocaine and crack, and heroin, respectively) were imputed in a manner similar to that described for cigarettes. However, some deviations from the process described for cigarettes applied to the AFU for these “other” substances as described below.

6.5.1.7.1 Response Propensity Step

The input to the weight adjustment model in the RP step for the AFU variables for other substances was the preliminary analysis weight. The RP step for AFU for other substances was very similar to the RP step for cigarette AFU (Section 6.5.1.4.1). For substances that included child measures, a response had to be provided for each substance for an individual to be considered a respondent for imputation purposes. Appendix D provides information on the covariates used in the models.

6.5.1.7.2 Prediction Step

The PRD step for AFU for other substances was also very similar to the analogous step for cigarettes (Section 6.5.1.4.2). For substances with child measures, only the parent measure was modeled. Modified versions of the 12-month frequency of use (where applicable) and AFU of previously imputed substances were used as covariates and were defined as follows:

- new12_i** = 0 if respondent *i* did not use the substance of interest in the past 12 months
- = 12-month frequency if respondent *i* used the substance of interest in the past 12 months
- AFU_i** = 0 if respondent *i* is not a lifetime user of the substance of interest
- = age at first use if respondent *i* is a lifetime user of the substance of interest

6.5.1.7.3 Hot-Deck Step

For cocaine (crack), more than one AFU variable was associated with a single predicted AFU. This led to a multivariate assignment of the imputed values. Respondents were asked for their

AFU of cocaine and were asked for their AFU of crack. This often led to complex and numerous logical constraints. These constraints used not only cocaine and crack ages at first use, when available, but also imputation-revised recencies and frequencies.

6.5.1.7.4 *Month-of-First-Use, Year-of-First-Use, and Day-of-First-Use Assignments*

The general principles described in Section 6.5.1.4.4 applied to the remaining substances with the following exceptions.

- For crack, the corresponding cocaine date of first use was assigned first. Then, in the setting of the earliest possible date for crack, the cocaine date of first use was used as an additional bound. This was done to ensure that the crack date of first use was never earlier than the cocaine date of first use.
- For cocaine, crack recency and frequency information was used to bound the latest possible date for the date of first use. For example, respondents with crack recency = 3 (i.e., lifetime but not past year user of crack) could not have first used cocaine within the past year, regardless of the cocaine recency value. The bound created using crack recency and frequency was calculated in exactly the same way as for cocaine recency and frequency information.

6.5.1.7.5 *Final Age and Date-of-First-Use Variables*

The final imputation-revised date-of-first-use variables for “other” variables, with the exception of smokeless tobacco, were named in the same manner as those for cigarettes: with the prefix IR, followed by a three-letter code identifying the substance and the final three letters identifying the measure (AGE = age at first use, MFU = month of first use, YFU = year of first use, DFU = day of first use). For smokeless tobacco, the six-letter code, SMKLSS, was used to identify the substance rather than a three-letter code. Additionally, TRY was used to identify the age-at-first-use measure. Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.5.1.8 30-Day Binge Drinking Frequency (Imputation Set 21)

For alcohol, an additional variable was defined that measured level of usage in the past month. In particular, the variable ALCBNG30D measured the binge drinking frequency or the number of days in the past 30 days on which the respondent had five or more drinks for males or four or more drinks for females on the same occasion. Since 30-day binge drinking is defined differently for males and females, the model building and final assignment of imputed values was performed separately within gender (male and female) and age group (12 to 17, 18 to 25, and 26 or older).

6.5.1.8.1 *Response Propensity Step*

The input to the weight adjustment model in the RP step for the 30-day binge drinking frequency variable was the preliminary analysis weight. The RP model was built using all past month users of alcohol whether they were binge drinkers or not. Item respondents for 30-day

binge drinking were defined as those past month alcohol users with a nonmissing value for ALCBNG30D. See Appendix D for information on the covariates used in the RP model for this variable.

6.5.1.8.2 *Prediction Step*

The response variable of interest in the 30-day binge drinking frequency model, prior to a normalizing transformation, was the proportion of the days in a month (30) on which a respondent had five or more drinks for males or four or more drinks for females on the same occasion. The range of values for the proportion was from 0 to 1, inclusive. Hence, to model 30-day binge drinking frequency of use, the following empirical logit transformation was computed for all respondents:

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

where Y_i was the observed 30-day binge drinking frequency for respondent i and N was 30, the total number of days in the month that the respondent could have binge drunk. This transformation was nearly equivalent to the standard logit transformation:

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

where P_i was defined as the proportion of days in the past month during which respondent i had five or more drinks for males or four or more drinks for females on the same occasion. The standard logit transformation was not used, because it was not defined for daily binge drinkers nor was it defined for nonbinge drinkers among past month users.¹³⁴ Using the adjusted weights, a linear univariate regression model was then fitted for the log-transformed variable Y_i within each gender and age group combination.

6.5.1.8.3 *Hot-Deck Step*

The hot-deck step for 30-day binge drinking frequency was straightforward and implemented separately for each of the age group and gender combinations. Two logical constraints were utilized. One required the donor's 30-day binge drinking frequency to be less than or equal to the recipient's 30-day alcohol use frequency. The other constraint utilized the recipients' responses to the question on the usual number of drinks of alcoholic beverages that they consumed on days when they used alcohol in the past 30 days. This constraint required the donors' 30-day binge drinking frequency to be strictly greater than zero if recipients reported usually consuming five or more drinks for males or four or more drinks for females on the days they consumed alcohol during the past 30 days. The state rank and delta likeness constraints were also applied. Details on the hot-deck step are available in Appendix E.

¹³⁴ See Cox and Snell (1989) for a discussion of the empirical logit transformation. If the respondent was a daily binge drinker of alcohol, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$, where Y was the observed 30-day binge drinking frequency and N was the total number of days that the respondent could have used (usually 30). If the proportion was zero, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[0.5/(N + 0.5)]$.

6.5.1.9 Marijuana Ways of Use (Imputation Sets 14 and 15)

Marijuana ways of use variables were imputed after marijuana recency and frequency (set 13) and before age at first marijuana use (set 16). These 16 variables are divided into two sets of 8 variables: set 14 includes past year ways of use and set 15 includes past month ways of use. Each set was imputed using the single RP/multiple PRD type of PMN imputation. As with other variables imputed using the PMN method, the data are divided and processed separately within three age groups, 12 to 17, 18 to 25, and 26 or older.

6.5.1.9.1 Response Propensity Step

Each set of eight edited ways of use variables is derived from one questionnaire item. If a respondent refused to answer or skipped the item, all eight variables would be missing. Therefore, only one RP model is fit for each age group, within each set of variables.

The input to the weight adjustment model in the RP step for the marijuana ways of use variables was the preliminary analysis weight. For the variables measuring ways of use in the past year, the domain included past year users of marijuana based on the imputation-revised marijuana recency variable (IRMJRC). To be classified as a respondent, all eight edited variables measuring ways of use in the past year had to be nonmissing (i.e., yes or no). For the variables measuring ways of use in the past month, the domain included past month users of marijuana based on the imputation-revised marijuana recency variable. To be classified as a respondent, all eight edited variables measuring ways of use in the past month had to be nonmissing. If any ways of use variables were logically assigned during editing, these values were propagated to the final imputed variable. However, these cases will still be classified as nonrespondents for modeling purposes. See Appendix D for information on the covariates used in the RP model for these variables.

6.5.1.9.2 Prediction Step

Across all age groups, the most common ways of use are smoking, vaping, eating or drinking, and dabbing waxes. Five PRD models for each age group were fit for each imputation set: one for each of these four most common ways of use, and a fifth to predict selection of one or more of the final four ways of use. These models were fit using weighted logistic regression. All five of the resulting predicted means were used to determine the donor neighborhoods in the hot-deck step.

6.5.1.9.3 Provisional Imputation Steps

After each PRD step, missing values were provisionally imputed using CSI. The resulting imputed values were used as predictors for subsequent models. The order for modeling and provisional imputation was (1) smoking, (2) vaping, (3) eating or drinking, (4) dabbing waxes, and (5) dissolving in mouth, lotion, pills, or other. Provisional imputation was only performed for the first four ways of use variables.

6.5.1.9.4 *Final Hot-Deck Steps*

After fitting the fifth PRD model in each set (past year and past month), all missing values were imputed in a single hot-deck step. Nonmissing ways of use variables that were logically assigned were used as likeness constraints on the donor pool by requiring the donor to match the recipient on these nonmissing values. All missing values within an imputation set were replaced by valid values from a single donor. Further details on the hot-deck step are available in Appendix E.

6.5.1.9.5 *Final Variables*

The imputed marijuana ways of use variables were named according to the corresponding edited variables, with the addition of the prefix IR. Past year marijuana ways of use variables were named IRMJYR followed by a description of the way of use: SMOKE, VAPE, DAB, EAT, MOUTH, SKIN, PILL, or OTHER. Past month marijuana ways of use variables were named similarly, using IRMJMON instead of IRMJYR. Each of the 16 imputed variables had a corresponding imputation indicator with the prefix II.

6.5.2 Variables Imputed Using the modPMN Methodology

The modPMN methodology is used for a subset of the substance use variables covering inhalants, hallucinogens (excluding hallucinogen child lifetime variables), methamphetamine, and prescription drugs. For details on the modPMN methodology, see Chapter 3. Section 6.5.2.1 describes how these variables were segregated into imputation sets. Section 6.5.2.2 describes how the variables were ordered within the imputation sets. Section 6.5.2.3 provides details on the steps of modPMN for categorical variables, and Section 6.5.2.4 provides details on the steps of modPMN for continuous variables. Section 6.5.2.5 describes the sequencing of steps for modPMN sets. Section 6.5.2.6.1 provides details on the imputation of date at first use of inhalants, hallucinogens, and methamphetamine, and Section 6.5.2.6.2 provides details on the age and date at first misuse of prescription drugs.

6.5.2.1 Imputation Sets for modPMN Substance Use Variables

Substance use variables imputed using the modPMN methodology are segregated into imputation sets by measure rather than by substance as is done under PMN methodology. However, not all substances for which a measure exists are included in a single imputation set. For example, child drugs (e.g., LSD, PCP, ECS) and specific prescription drugs (e.g., OxyContin®, benzodiazepines, and fentanyl products) are imputed in separate sets following the imputation of the parent measure (e.g., hallucinogens) or the more general drug category (e.g., pain relievers for OxyContin® and fentanyl products) so that the domain for these measures can be fully defined. In addition, initiation data for the prescription drugs are imputed separately from inhalants, hallucinogens, and methamphetamine due to the difference in domain and imputation methodology (Section 6.5.2.6). CBD or hemp and nicotine vaping are imputed in separate sets because they have different predecessors, requiring different placement in the processing timeline. [Table 6.11](#) shows the 30 imputation sets covered in this section. Note that imputation sets 1 through 22 are covered in Section 6.5.1. Sets 25 through 27, 29 through 31, 37 through 40, and 42 through 44 can be run in any order (or in parallel). However, sets 25 through 27

cannot be run until after set 24 has been run, sets 29 through 31 cannot be run until after set 28 and the corresponding recency of any use set has been completed, sets 37 through 40 cannot be run until after set 36 is completed, and sets 42 through 44 cannot be run until after set 41 is completed. Set 46 can be run after set 1 and before sets 23 through 45. However, Set 47 cannot be run until after Set 11. Set 51 cannot be run until after set 23, as well as the imputation of lifetime vaping of flavoring, as described in Section 8.6.6.

Table 6.11 Substances and Substance Use Measures, modPMN Imputation Sets

Substance	Substance Use Measures							Age at First Use/ Misuse	Date of First Use/ Misuse
	Life-time Usage	Recency of Any Use	Recency of Use/ Misuse	12-Month Frequency of Use	30-Day Frequency of Use/ Misuse	Past Year Initiation			
Inhalants									Set 42
Hallucinogens	Set 23	N/A		Set 33		N/A	Set 36		Set 43
Methamphetamine									Set 44
Stimulants			Set 28		Set 34				Set 37
Pain Relievers									Set 38
Tranquilizers		Set 24				Set 35			Set 39
Sedatives	N/A								Set 40
OxyContin®		Set 25 ¹	Set 29 ¹						
Benzodiazepines		Set 26 ¹	Set 30 ¹	N/A			N/A	N/A	
Fentanyl Products		Set 27 ¹	Set 31 ¹						
LSD, PCP, Ecstasy ²					N/A			Set 41	Set 45
Ketamine, DMT/AMT/ "Foxy," <i>Salvia divinorum</i> ²	DM	N/A	Set 32			N/A		N/A	N/A
Nicotine Vaping	Set 46		Set 47		Set 48		Set 49	Set 50	
CBD or Hemp	Set 51		Set 52		N/A		N/A	N/A	

DM = different method; modPMN = modified predictive mean neighborhood; N/A = not applicable; PMN = predictive mean neighborhood.

¹ Recency of Any Use and Recency of Misuse for OxyContin®, benzodiazepines, and fentanyl products include two levels because only use in the past 12 months is asked in the survey. Recency of Any Use includes levels for past year any use and did not use in the past year. Recency of Misuse includes levels for past year misuse and did not misuse in the past year.

² Hallucinogen child drugs recency/frequency and age at first use are imputed using modPMN. Lifetime use for hallucinogen child drugs are imputed using PMN methodology. See [Table 6.7](#) for the PMN imputation sets.

6.5.2.2 Imputation Order within Imputation Sets

As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rates. For sets that included both prescription drug and nonprescription drug measures (i.e., imputation sets 28 and 34), inhalants, hallucinogens, and methamphetamine preceded all prescription drugs. The ordering for 2023 is shown in [Table 6.12](#).

Table 6.12 Sequencing of Substances within Imputation Sets, 2023

Imputation Set	Order
Lifetime Substance Use	Methamphetamine, Inhalants, Hallucinogens
Recency of Any Use	Stimulants, Tranquilizers, Sedatives, Pain Relievers
Recency of Use/Misuse	Methamphetamine, Hallucinogens, Inhalants, Stimulants, Tranquilizers, Sedatives, Pain Relievers
Hallucinogen Child Drug Recency of Use	Ecstasy, LSD, PCP, Salvia divinorum, Ketamine, DMT/AMT/"Foxy"
12-Month Frequency of Use	Methamphetamine, Hallucinogens, Inhalants
30-Day Frequency of Use	Methamphetamine, Inhalants, Hallucinogens, Stimulants, Tranquilizers, Sedatives, Pain Relievers
Age at First Use	Methamphetamine, Hallucinogens, Inhalants
Hallucinogen Child Drugs Age at First Use	Ecstasy, LSD, PCP
Past Year Initiation	Stimulants, Tranquilizers, Sedatives, Pain Relievers

6.5.2.3 Categorical Variables

This section provides details on the three steps of modPMN for categorical variables as described in Sections 3.6.1 through 3.6.3. Due to the number of imputation sets, set 28 is explained in detail in Sections 6.5.2.3.1 through 6.5.2.3.4, and the rest of the categorical variable sets are described in Section 6.5.2.3.5 as deviations from these procedures.

Categorical variables include lifetime usage (sets 23, 46, and 51), recency of any use (sets 24 through 27), recency of (mis)use (sets 28 through 32, 47, and 52), and past year initiation (set 35).

6.5.2.3.1 Recency of (Mis)Use, Set 28: Response Propensity Adjustment

This section provides specific information about the RP steps for the recency of (mis)use variables (imputation set 28). Section 3.6.1 provides general information about the RP adjustment step of modPMN. One RP model was fit for each substance in the set for each cycle, for a total of 14 RP models for this set. The four key inputs to the RP adjustment step, as laid out in Section 3.3.1, are described below.

- **Analysis Weight:** The input to all RP models was the preliminary analysis weight.
- **Domain:** For inhalants, hallucinogens, and methamphetamine, the domain includes lifetime users of the substance based on the imputation-revised lifetime use indicator. For prescription drugs, the domain includes lifetime users of the drugs (IRXXXANYREC=1, 2) based on the imputation-revised any use recency variable. See [Table 6.13](#) for domains for all other sets.
- **Item Response Indicator:** If a valid response was provided for the recency of (mis)use measure for the substance, the individual was deemed an item respondent.
- **Covariates:** See Appendix D for information on the covariates used in the RP models.

6.5.2.3.2 *Recency of (Mis)Use, Set 28: Prediction Modeling*

This section provides specific information about the PRD modeling step for the recency of (mis)use set 28. Section 3.6.2 provides general information about the PRD modeling step of modPMN. Each of the 14 RP models segregated the domain members (including both respondents and nonrespondents) into five bins of roughly equal size, and separate PRD models were fit within each of the bins. If one or more bins had fewer than 10 respondents, and one of the bins with fewer than 10 respondents also included at least 1 nonrespondent, the five bins were collapsed, and a single PRD model was fit using respondents from all five RP bins combined. This resulted in a maximum of 70 total PRD models for imputation set 28 (7 substances × 5 bins × 2 cycles).

For each of the PRD models, the probability of selecting each recency of (mis)use category was modeled within each bin using polytomous logistic regression. SUDAAN'S MULTILOG procedure was used to estimate the parameters from the appropriate logistic model, which yielded predicted probabilities for each of the recency categories for both item respondents and item nonrespondents. For inhalants, hallucinogens, and methamphetamine, the three recency categories were the following:

- past month use;
- past year, not past month use; and
- lifetime, not past year use.

For prescription drugs, the four recency categories were the following:

- past month misuse;
- past year, not past month misuse;
- lifetime, not past year misuse; and
- never misused.

These predicted probabilities were used in the subsequent step to impute a recency category of (mis)use. See Appendix D for information on the covariates used in the PRD models.

6.5.2.3.3 *Recency of (Mis)Use, Set 28: Collocated Stochastic Imputation*

This section provides specific information about the CSI step for set 28. Section 3.6.3 provides general information about the CSI step. Each of the PRD models is followed by an imputation step where the predicted probabilities from the previous step, along with a random number, is used to impute a recency category of (mis)use for the given substance for each item nonrespondent.

In the circumstance of partial information, the predictive mean vector is adjusted to prevent inconsistent imputation. For example, if a respondent's edited recency of inhalant use is past year where past month use versus past year not past month use is unknown, the predicted probability associated with lifetime not past year use, $P(\text{Lifetime but not past year use of inhalants} \mid \text{lifetime use of inhalants})$, is set to 0. The remaining probabilities corresponding to past month use and past year not past month use are then rescaled to sum to 1 prior to

imputation. For more information on the restrictions based on lifetime use (for inhalants, hallucinogens, and methamphetamine) and recency of any use (for prescription drugs), see Appendix E.

6.5.2.3.4 *Recency of (Mis)Use, Set 28: Final Variables*

The final imputation-revised recency-of-use variables for hallucinogens (HALLUC), inhalants (INHAL), and methamphetamine (METHAM) were identified with the prefix IR, followed by the substance identifier and a final "REC" to identify the measure. For the prescription drugs, the final imputation-revised recency of misuse variables were identified with the prefix IR, followed by an eight-letter identifier, where a three-letter code identified the drug (PNR, SED, STM, TRQ) and the final five letters identified the measure (NMREC). Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

6.5.2.3.5 *"Other" Categorical ModPMN Variables*

Imputation for sets 23 through 27, 29 through 32, 35, 46, 47, 51, and 52 in [Table 6.11](#) was done in a manner similar to that described above for set 28. The following deviations from the process described for set 28 applied to these "other" categorical modPMN variables.

6.5.2.3.5.1 *Response Propensity Step*

The overall process for the RP step was similar to the process described for set 28; however, the number of models fit and the inputs to the RP step varied from set to set. The total number of RP models fit for each set can be found in [Table 6.13](#). Following are the four key inputs to the RP adjustment step.

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain for each set is detailed in [Table 6.13](#).
- **Item Response Indicator:** If a valid response was provided for the given measure for the substance, the individual was deemed an item respondent.
- **Covariates:** Appendix D provides information on the covariates used in the RP models.

Table 6.13 Summary of Response Propensity Models for Categorical modPMN Variables

Set	Number of Response Propensity Models	Domain
23: Lifetime Usage	6	All respondents
24: Recency of Any Use	8	All respondents
25: Recency of Any Use, OxyContin®	1	Past year users of pain relievers, based on IRPNRANYREC
26: Recency of Any Use, Benzodiazepines	1	Past year users of sedatives or tranquilizers, based on IRSEDANYREC and IRTRQANYREC
27: Recency of Any Use, Fentanyl Products	1	Past year users of pain relievers, based on IRPNRANYREC

Table 6.13 Summary of Response Propensity Models for Categorical modPMN Variables (continued)

Set	Number of Response Propensity Models	Domain
28: Recency of (Mis)Use	14	Nonprescription substances: Lifetime users, based on imputation-revised lifetime use Prescription drugs: Lifetime users, based on IRXXXANYREC
29: Recency of Misuse, OxyContin®	1	Past year misusers of pain relievers, based on IRPNRNMREC and past year users of OxyContin® based on IROXCNANYR
30: Recency of Misuse, Benzodiazepines	1	Past year misusers of sedatives or tranquilizers, based on IRSEDNMREC and IRTQNMREC, and past year users of benzodiazepines, based on IRBZOANYR
31: Recency of Misuse, Fentanyl Products	1	Past year misusers of pain relievers, based on IRPNRNMREC, and past year users of fentanyl products, based on IRFENTANYR
32: Recency of Use, Hallucinogen Child Drugs	12	Lifetime users of hallucinogens, based on IRHALLUCREC
35: Past Year Initiation	8	Past year misusers, based on IRXXXNMREC
46: Lifetime Use, Nicotine Vaping	1	All respondents
47: Recency of Use, Nicotine Vaping	1	Lifetime users, based on IRNICVAPREC
51: Lifetime Use, CBD or Hemp	1	All respondents
52: Recency of Use, CBD or Hemp	1	Lifetime users, based on IRCBDHMPREC

modPMN = modified predictive mean neighborhood.

6.5.2.3.5.2 Prediction Modeling Step

Section 6.5.2.3.2 described the SUDAAN method used to impute a multilevel variable. Other sets that model multilevel variables include sets 24, 32, 47, and 52. The other categorical variable modPMN sets model dichotomous variables. The PRD model for these sets utilized dichotomous logistic regression through SUDAAN's LOGISTIC procedure (called RLOGIST in SAS-callable SUDAAN). The levels for outcome variables for each set are listed in [Table 6.14](#). See Appendix D for information on the covariates for the PRD models.

Table 6.14 Outcomes of Prediction Modeling

Set	Outcome Variable Levels
23: Lifetime Usage	1: Lifetime Use 2: Never Used
24: Recency of Any Use	1: Past Year Use 2: Lifetime, Not Past Year Use 3: Never Used
25: Recency of Any Use, OxyContin®	1: Past Year Use 2: No Past Year Use

Table 6.14 Outcomes of Prediction Modeling (continued)

Set	Outcome Variable Levels
26: Recency of Any Use, Benzodiazepines	1: Past Year Use 2: No Past Year Use
27: Recency of Any Use, Fentanyl Products	1: Past Year Use 2: No Past Year Use
28: Recency of (Mis)Use	<i>Nonprescription Substances</i> 1: Past Month Use 2: Past Year, Not Past Month Use 3: Lifetime, Not Past Year Use <i>Prescription Drugs</i> 1: Past Month Misuse 2: Past Year, Not Past Month Misuse 3: Lifetime, Not Past Year Misuse 4: Never Misused
29: Recency of Misuse, OxyContin®	1: Past Year Misuse 2: No Past Year Misuse
30: Recency of Misuse, Benzodiazepines	1: Past Year Misuse 2: No Past Year Misuse
31: Recency of Misuse, Fentanyl Products	1: Past Year Misuse 2: No Past Year Misuse
32: Recency of Use, Hallucinogen Child Drugs	1: Past Month Use 2: Past Year, Not Past Month Use 3: Lifetime, Not Past Year Use
35: Past Year Initiation	1: Past Year Misuse Initiate 2: Not a Past Year Misuse Initiate
46: Lifetime Use, Nicotine Vaping	1: Lifetime Use 2: Never Used
47: Recency of Use, Nicotine Vaping	1: Past Month Use 2: Past Year, Not Past Month Use 3: Past 3 Years, Not Past Year Use 4: Lifetime, Not Past 3 Years Use
51: Lifetime Use, CBD or Hemp	1: Lifetime Use 2: Never Used
52: Recency of Use, CBD or Hemp	1: Past Month Use 2: Past Year, Not Past Month Use 3: Past 3 Years, Not Past Year Use 4: Lifetime, Not Past 3 Years Use

6.5.2.3.5.3 CSI Step

There are no deviations from the CSI procedure described in Section 6.5.2.3.3. Partial information is not available for the edited variables in sets 23, 25 through 27, and 52; therefore, no adjustments to the predictive mean vectors are made in these sets. In the other sets, partial information from edited variables and, in some cases, previously imputed substance use values, are also used to adjust the predictive mean vectors. Additionally, in set 32, edited hallucinogen age at first use is used to adjust the hallucinogen child drug recency-of-use predictive mean vectors to ensure consistency between parent and child recency and AFU. For more information on these restrictions, see Appendix E.

6.5.2.3.5.4 Final Variables

No final imputation-revised variables indicating lifetime usage (sets 23, 46, and 51) alone were created because this information was recoded in the final imputation-revised recency of (mis)use variables. Imputation indicators also were not created.

The final imputation-revised variables for the other sets followed a similar pattern as those for set 28: the IR prefix followed by a substance abbreviation and ending with an indicator of the specific measure. [Table 6.15](#) provides the final imputation-revised variable format for each set. Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

Table 6.15 Final Imputation-Revised Categorical modPMN Variables

Set	IR Variables ¹
23: Lifetime Usage	N/A
24: Recency of Any Use	IRXXXANYREC
25: Recency of Any Use, OxyContin®	IROXCANYR
26: Recency of Any Use, Benzodiazepines	IRBZOANYR
27: Recency of Any Use, Fentanyl Products	IRFENTANYR
28: Recency of (Mis)Use	Nonprescription Substances: IRXXXREC Prescription Drugs: IRXXXNMREC
29: Recency of Misuse, OxyContin®	IROXCNNMYR
30: Recency of Misuse, Benzodiazepines	IRBZONMYR
31: Recency of Misuse, Fentanyl Products	IRFENTNMYR
32: Recency of Use, Hallucinogen Child Drugs	LSD and PCP: IRXXXRC Others: IRXXXREC
35: Past Year Initiation	IRXXXNMINIT
47: Recency of Use, Nicotine Vaping	IRNICVAPREC
52: Recency of Use, CBD or Hemp	IRCBDHMPREC

modPMN = modified predictive mean neighborhood; N/A = not applicable.

¹ XXX represents the substance abbreviation. Sometimes it may be longer than three characters (e.g., HALLUC for hallucinogens).

6.5.2.4 Continuous Variables

This section provides details on the three steps of modPMN for continuous variables as described in Sections 3.6.1 through 3.6.3. Set 33 is explained in detail in Sections 6.5.2.4.1 through 6.5.2.4.4, and the rest of the continuous variable sets are described in Section 6.5.2.4.5 as deviations from these procedures.

Continuous variables include 12-month frequency of use (set 33), 30-day frequency of (mis)use (sets 34 and 48), and age at first use (sets 36, 41, and 49).

6.5.2.4.1 12-Month Frequency of Use, Set 33: Response Propensity Adjustment

This section provides specific information about the RP steps for the 12-month frequency-of-use set 33. Section 3.6.1 provides general information about the RP adjustment step of modPMN. One RP model was fit for each substance in the set for each cycle, for a total of six RP models for this set. The four key inputs to the RP adjustment step, as laid out in Section 3.3.1, are described below.

- **Analysis Weight:** The input to all RP models was the preliminary analysis weight.
- **Domain:** The domain includes past year users of the substance based on the imputation-revised recency of use variable, IRXXXREC.
- **Item Response Indicator:** If a valid response was provided for the 12-month frequency-of-use measure for the substance, the individual was deemed an item respondent.
- **Covariates:** See Appendix D for information on the covariates used in the RP models.

6.5.2.4.2 12-Month Frequency of Use, Set 33: Prediction Modeling

This section provides specific information about the PRD modeling step for the 12-month frequency-of-use set 33. Section 3.6.2 provides general information about the PRD modeling step of modPMN. Each of the six RP models segregated the domain members (including both respondents and nonrespondents) into five bins of roughly equal size, and separate PRD models were fit within each of the bins. If one or more bins had fewer than 10 respondents, and one of the bins with fewer than 10 respondents also included at least 1 nonrespondent, the five bins were collapsed, and a single PRD model was fit using respondents from all five RP bins combined. This resulted in a maximum of 30 total PRD models for imputation set 33 (3 substances × 5 bins × 2 cycles).

The response variable of interest in the 12-month frequency-of-use models for most respondents, prior to a normalizing transformation, was the proportion of the days in a full year (365.25) on which a respondent used a particular substance. Some respondents, however, started using the substance within the past year. In these instances, this additional information was used to reduce the number of days the respondent could have used the substance in the past year. The range of values of this proportion of days used is from (greater than) 0 to 1. Therefore, in order to model 12-month frequency of use, an empirical logit transformation was

computed for all respondents. Then, a linear univariate regression model using SUDAAN software was fit. For details on the transformation and reasons for use, see Section 6.5.1.6.2.2.

The PRD models produced predicted means for both respondents and nonrespondents that were a logit of the proportion of the year used. These predicted means were back-transformed into a proportion for use as the variables from which the neighborhoods were created. See Appendix D for information on the covariates used in the PRD models.

6.5.2.4.3 12-Month Frequency of Use, Set 33: Hot-Deck

This section provides specific information about the hot-deck imputation step for set 33. As discussed in Section 3.6.3, the hot-deck step for modPMN almost always employs univariate matching and univariate assignment.

As discussed in Section 6.2.4.5, edit rules were used to create bounds for data missing the 12-month frequency for inhalants and methamphetamine. Specifically, the minimum number of days a recipient could have used the substance in the past 12 months was defined as follows:

- If the recipient was known to have initiated more than 30 days ago but within the past year, was a past month user of the substance, and the 30-day frequency was known, then the minimum number of days was the 30-day frequency + 1.
- If the recipient was known to have initiated more than 30 days ago but within the past year, was a past month user of the substance, and the estimated 30-day frequency was known, then the minimum number of days was the lower bound of the estimated interval + 1.
- If the recipient was known to have initiated more than 30 days ago but within the past year, was a past month user of the substance, and the 30-day frequency was unknown, then the minimum number of days was 2.
- If the recipient's initiation was unknown or the recipient was known to have initiated within the past 30 days or more than a year ago, was a past month user of the substance, and the 30-day frequency was known, then the minimum number of days was the 30-day frequency.
- If the recipient's initiation was unknown or the recipient was known to have initiated within the past 30 days or more than a year ago, was a past month user of the substance, and the estimated 30-day frequency was known, then the minimum number of days was the lower bound of the estimated interval.
- If the recipient's initiation was unknown or the recipient was known to have initiated within the past 30 days or more than a year ago, was a past month user of the substance, and the 30-day frequency was unknown, then the minimum number of days was 1.
- If the recipient was a past year but not past month user, then the minimum number of days was 1.

The maximum number of days a recipient could have used the substance in the past 12 months was defined as follows:

- If the number of days between the earliest possible initiation date of the recipient and the interview date (MAXDAYS) was less than or equal to 30, and the 30-day frequency was known, then the maximum number of days was the 30-day frequency.
- If MAXDAYS was less than or equal to 30, and the estimated 30-day frequency was known, then the maximum number of days was the upper bound of the estimated interval.
- If MAXDAYS was less than or equal to 30, and the 30-day frequency was unknown, then the maximum number of days was MAXDAYS.
- If MAXDAYS was more than 30 but less than 365, the recipient was a past month user, and the 30-day frequency was known, then the maximum number of days was $\text{MAXDAYS} - 30 + \text{the 30-day frequency}$.
- If MAXDAYS was more than 30 but less than 365, the recipient was a past month user, and the estimated 30-day frequency was known, then the maximum number of days was $\text{MAXDAYS} - 30 + \text{the upper bound of the estimated interval}$.
- If MAXDAYS was more than 30 but less than 365, the recipient was a past month user, and the 30-day frequency was unknown, then the maximum number of days was MAXDAYS.
- If MAXDAYS was more than 30 but less than 365, and the recipient was a past year but not past month user, then the maximum number of days was $\text{MAXDAYS} - 30$.
- If MAXDAYS was greater than or equal to 365 days, the recipient was a past month user, and the 30-day frequency was known, then the maximum number of days was $335 + \text{the 30-day frequency}$.
- If MAXDAYS was greater than or equal to 365 days, the recipient was a past month user, and the estimated 30-day frequency was known, then the maximum number of days was $335 + \text{the upper bound of the estimated interval}$.
- If MAXDAYS was greater than or equal to 365 days, the recipient was a past month user, and the 30-day frequency was unknown, then the maximum number of days was 365.
- If MAXDAYS was greater than or equal to 365 days, and the recipient was a past year but not a past month user, then the maximum number of days was 335.

These bounds were used to form a single logical constraint for inhalants and methamphetamine. Logical constraints for hallucinogens involved the interview date, initiation data (AFU, MFU, and YFU), birthday, imputation-revised recency of use, and 30-day frequency of use. In addition to the delta constraint, the likeness constraints required recipients and donors to have the same recency of use and be in the same state-rank group based on level of past month usage. Additionally, recipients and donors were required to be in the same RP bin. If no donors were available within these constraints, then they were loosened in the following order: (1) the delta constraint was removed, (2) donors and recipients were no longer required to be from states with similar usage levels, and (3) donors and recipients were no longer required to have the same recency of use. If donors were still not available, the three

constraints listed above were reapplied and the constraint requiring donors and recipients to be in the same RP bin was dropped. If no donors could be found after dropping the bin constraint, the likeness constraints that had been reapplied were once again loosened in the same order as described previously until a nonempty neighborhood of donors was obtained. A complete list of likeness and logical constraints applied to the 12-month frequency-of-use imputations can be found in Appendix E.

6.5.2.4.4 12-Month Frequency of Use, Set 33: Final Variables

The final imputation-revised 12-month frequency-of-use variables for hallucinogens (HALLUC), inhalants (INHAL), and methamphetamine (METHAM) were identified with the prefix IR, followed by the substance identifier and a final "YFQ" to identify the measure. Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

6.5.2.4.5 "Other" Continuous ModPMN Variables

Imputation for sets 34, 36, 41, 48, and 49 in [Table 6.11](#) was completed in a manner similar to that described above for set 33. The following deviations applied to these "other" continuous modPMN variables. For set 34, the preliminary setup program is described in Section 6.5.2.4.5.1. Additionally, for the prescription drug variables in set 34, the edit rules described in Section 6.3.13 were run on the imputation-revised data to ensure that no unallowable inconsistencies were present.

6.5.2.4.5.1 30-Day Frequency of (Mis)Use, Sets 34 and 48: Preliminary Setup

For 30-day frequency of (mis)use, several steps not involving any modeling or stochasticity are handled in a preliminary setup program as follows:

- Bounds are created for respondents missing exact 30-day frequency:
 - For prescription drugs, the lower bound is set to 1, and the upper bound is created using the edit rules as described in Section 6.3.13.
 - For hallucinogens, both the lower and upper bounds are created based on the imputation-revised 12-month frequency of use. If the imputation-revised 12-month frequency of use is less than 30, then it forms the upper bound on the 30-day frequency of use. If the imputation-revised 12-month frequency of use is greater than 335, then the difference between the frequency and 335 is the lower bound on the 30-day frequency of use. For inhalants and methamphetamine, bounds are created based on the maximum number of days the respondent could have used in the past year given the known incidence data (MAXDAYS) and the imputation-revised 12-month frequency of use.
- The lower bound is defined as the maximum of the following candidate lower bounds:
 - If the imputation-revised 12-month frequency of use is greater than 30 and is greater than MAXDAYS – 30, then one of the candidate lower bounds is the

- imputation-revised 12-month frequency minus the number of days the respondent used prior to the past 30 days (MAXDAYS – 30).
- If the respondent definitely initiated in the past 30 days, then one of the candidate lower bounds is the imputation-revised 12-month frequency.
 - For all respondents with missing exact 30-day frequency, one of the candidate lower bounds is 1.
- The upper bound is defined as the minimum of the following candidate upper bounds:
 - If the respondent initiated on the interval (1 year ago, 30 days ago), then one of the candidate upper bounds is the imputation-revised 12-month frequency minus 1.
 - If the respondent did not initiate on the interval (1 year ago, 30 days ago), then one of the candidate upper bounds is the imputation-revised 12-month frequency.
 - For all respondents with missing exact 30-day frequency, one of the candidate upper bounds is 30.
 - For nicotine vaping, the lower bound is set to 1, and the upper bound is created based on the maximum number of days the respondent could have used in the past year given the known incidence data.
 - When a respondent is missing the exact 30-day frequency of (mis)use but not missing the estimated 30-day frequency of (mis)use, the final edited 30-day frequency of (mis)use was assigned to the midpoint of the interval associated with the response category. For example, respondents selecting “10 to 19 days” as a response to the estimated 30-day frequency-of-use question were assigned a final edited 30-day frequency of use of 14.5. Sometimes the bounds from the preceding step were used to shorten the interval before assigning the midpoint. For example, if the estimated 30-day frequency of (mis)use was “10 to 19 days,” but the upper bound for the 30-day frequency was 15, then the final edited 30-day frequency of use was 12.5 instead of 14.5.
 - The imputation indicator variables are defined. See Section 6.5.2.4.5.5 for more information on the final imputation-revised and imputation indicator variables.
 - The item response indicator used in the RP and PRD steps along with indicators used in the hot-deck step are defined.

6.5.2.4.5.2 Response Propensity Step

The overall process for the RP step was similar for the sets; however, the number of models fit and the inputs to the RP step varied from set to set. The total number of RP models fit for each set can be found in [Table 6.16](#). For each of the four key inputs to the RP adjustment step, details on any deviations are described below.

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain for each set is detailed in [Table 6.16](#).

- **Item Response Indicator:** If a valid response was provided for the given measure for the substance of interest (i.e., item response was determined separately for each individual substance rather than across all substances within a given imputation set), the individual was deemed an item respondent. For sets 36, 41, and 49, respondents who reported an age at first use of 1 or 2 were treated as item nonrespondents in the RP and PRD steps because of the implausibility of such a young age at first use. In the hot-deck step, their response was left unchanged, but they were not allowed to be donors.
- **Covariates:** See Appendix D for information on the covariates used in the RP models.

Table 6.16 Summary of Response Propensity Models for Continuous modPMN Variables

Set	Number of Response Propensity Models	Domain
33: 12-Month Frequency of Use	6	Past year users, based on imputation-revised recency of use
34: 30-Day Frequency of (Mis)Use	14	Past month users, based on imputation-revised recency of (mis)use
36: Age at First Use	6	Lifetime users, based on imputation-revised recency of use
41: Age at First Use, Hallucinogen Child Drugs	6	Lifetime hallucinogen child drug users, based on imputation-revised recency of use
48: 30-Day Frequency of Use, Nicotine Vaping	1	Past month users, based on IRNICVAPREC
49: Age at First Use, Nicotine Vaping	1	Lifetime users, based on IRNICVAPREC

modPMN = modified predictive mean neighborhood.

6.5.2.4.5.3 Prediction Modeling Step

As detailed in Section 6.5.2.4.2, each of the RP models segregated the domain members (including both respondents and nonrespondents) into five bins of roughly equal size, and separate PRD models were fit within each of the bins. If one or more bins had fewer than 10 respondents, and one of the bins with fewer than 10 respondents also included at least 1 nonrespondent, the bins were collapsed, and a single PRD model was fit using respondents from all five RP bins combined. Descriptions of the variable of interest for each set are detailed below.

- **30-Day Frequency of (Mis)Use, Sets 34 and 48:** The response variable of interest in the 30-day frequency-of-(mis)use models, prior to a normalizing transformation, was the proportion of the days in the past month (30 days) on which a respondent (mis)used the substance of interest. The range of values for the proportion was from (greater than) 0 to 1. Therefore, to model 30-day frequency of (mis)use, an empirical logit transformation was computed for all respondents. See Section 6.5.1.3.6 for details on the transformation.

- **Age at First Use, Sets 36, 41, and 49:** The response variable of interest in the AFU model, prior to a normalizing transformation, was the AFU as a proportion of the current age. The numerator in this proportion was an integer representing AFU. However, because this integer was in fact a truncated version of the real AFU, the value was made continuous by adding a random component between 0 and 1. The denominator in the proportion was the total age calculated from the respondent's birth date and interview date. The range of values for the proportion was from (greater than) 0 to 1. Therefore, to model AFU, an empirical logit transformation was computed for all respondents. See Section 6.5.1.4.2 for details on the transformation and calculation of the AFU as a proportion of the current age.

For all sets included in [Table 6.16](#), a linear univariate regression model using SUDAAN software was fit, and the predicted means were calculated for both respondents and nonrespondents. These predicted means from the models were then back-transformed for use in the hot-deck step. See Appendix D for information on the covariates used in the PRD models.

6.5.2.4.5.4 Hot-Deck Step

Because all hot-deck steps for modPMN employ univariate matching and univariate assignment, the only deviations are found in the logical and likeness constraints applied. Below are details for sets 34, 36, 41, 48 and 49.

- **30-Day Frequency of (Mis)Use, Sets 34 and 48:** Due to small domain size for 30-day frequency of (mis)use and to increase the likelihood of keeping the delta constraint, the likeness constraint requiring donors and recipients to be in the same RP bin was not applied. For a complete list of logical and likeness constraints used, see Appendix E.
- **Age at First Use, Sets 36, 41, and 49:** Appendix E provides a complete list of likeness and logical constraints applied to the AFU imputations. The likeness constraints for AFU were more stringent than those for the other substance use measures, often resulting in the need to loosen the constraints.

6.5.2.4.5.5 Final Variables

The final imputation-revised variables for the other sets followed a similar pattern as those for set 33: the IR prefix followed by a substance abbreviation and ending with an indicator of the specific measure. [Table 6.17](#) provides the final imputation-revised variable format for each set. Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

Table 6.17 Final Imputation-Revised Continuous modPMN Variables

Set	IR Variables ¹
33: 12-Month Frequency of Use	IRXXXYFQ
34: 30-Day Frequency of (Mis)Use	Nonprescription Substances: IRXXX30N Prescription Drugs: IRXXXNM30FQ
36: Age at First Use	IRXXXAGE
41: Age at First Use, Hallucinogen Child Drugs	PCP and LSD: IRXXXAGE Ecstasy: IRECSTMOAGE
48: 30-Day Frequency of Use, Nicotine Vaping	IRNICVAP30N
49: Age at First Use, Nicotine Vaping	IRNICVAPAGE

modPMN = modified predictive mean neighborhood.

¹ XXX represents the substance abbreviation. Sometimes it may be longer than three characters (e.g., HALLUC for hallucinogens).

6.5.2.5 Sequencing of Steps for ModPMN Sets

This section provides the sequencing of steps for the modPMN sets 23 through 36 and 41. General information about the sequencing of steps for modPMN is available in Sections 3.7.2 and 3.7.3.

Imputation sets 23, 33, and 36 include three variables, so there are six 12-step blocks of programs included. The first of the six is for the first variable within the first cycle; the second is for the second variable within the first cycle; the third is for the third variable within the first cycle; the fourth is for the first variable within the second cycle; the fifth is for the second variable within the second cycle; and the sixth and last is for the third variable within the second cycle. Each 12-step block consists of an RP step where the domain members are segregated into five bins, followed by a PRD modeling step within each bin, followed by a collocated stochastic imputation step, and finally followed by a “wrap-up” program that merges the five datasets into one. ¹³⁵

Imputation sets 24 and 35 include four variables, imputation set 32 includes six variables, and imputation sets 28 and 34 include seven variables. The algorithm is identical to the one used for imputation sets 23, 33, and 36, except there are a different number of 12-step blocks per cycle.

Imputation sets 25 through 27, 29 through 31, 46 through 49, 51, and 52 include a single variable that eliminates the need for cycling. Therefore, each of these sets has a single 12-step block of programs.

¹³⁵ For some variables, the imputation step of modPMN is completed in a single SAS program rather than a separate program for each bin, resulting in fewer than 12 programs being used. However, in these situations, the bin restrictions are still included in the imputation step. That is, CSI is still performed separately within each bin for categorical variables, and a likeness constraint is used in the hot-deck step to restrict the donor pool based on the bin of the recipient for continuous variables.

6.2.5.6 Random Imputation, Sets 37 through 40, 42 through 45, and 50

This section provides specific information about the random imputation of the date of first use for inhalants, hallucinogens, methamphetamine, and nicotine vaping (sets 42 through 45 and 50) along with the age and date of first misuse of the prescription drugs (sets 37 through 40).

6.2.5.6.1 *Date-of-First-Use Assignments*

After the AFU imputations, all lifetime users of the substance of interest had nonmissing AFU values.¹³⁶ Using this AFU, users were assigned values for MFU, YFU, and DFU when these values were missing. Recent initiates, or those respondents whose AFU was within 1 year of their current age, were asked for their MFU and YFU. The DFU was not collected in the questionnaire and was missing for all respondents. The MFU, YFU, and DFU data contained four patterns of missingness:

- Pattern 1:** Recent initiates: missing day of first use only;
- Pattern 2:** Recent initiates: missing month/day of first use;
- Pattern 3:** Recent initiates: missing year/month/day of first use; and
- Pattern 4:** Less recent initiates: missing year/month/day of first use.

For each missingness pattern, upper and lower bounds on the date of first use were determined. Once the earliest and latest possible dates of first use were determined, a day was randomly selected from this interval. The imputation-revised month/day/year values were then extracted from this date of first use. For more information on the missingness patterns, see Sections 6.5.1.4.4.1 through 6.5.1.4.4.4.

There were no deviations from this general method when assigning date-of-first nicotine vaping use. For inhalants and methamphetamine, the edit rules described in Sections 6.2.4.5 and 6.5.2.4.3 were used to identify unallowable intervals for the date of first use. Specifically, if the 12-month frequency was equal to the 30-day frequency, then the randomly selected date of first use could not be on the interval (1 year ago, 30 days ago). If the randomly selected date of first use was on this interval, then up to 50 attempts were made to randomly select a date of first use outside of this interval. This approach was taken because it is sometimes difficult to determine the exact interval of valid dates of first use for data with 12-month frequency equal to 30-day frequency.

In fact, it is even possible for the set of valid dates of first use to form two intervals instead of one. Consider the following data:

- BRTHDATE = 3/26/1993
- INTDATE = 4/10/2023
- AFU = 29

¹³⁶ For prescription drugs, a date of first misuse was randomly selected and used to calculate the respondent's age at first misuse. For more details on age and date of first misuse, see Section 6.5.2.6.2.

- 12-month frequency = 30-day frequency
- Missing year, month, and date of first use

The set of valid dates of first use includes dates on the interval [3/26/2022, 4/10/2022] and dates on the interval [3/12/2023, 3/25/2023].

For hallucinogens, the child recency and frequency information was used to bound the latest possible date for the date of first use. For example, respondents with LSD recency = 3 (i.e., lifetime but not past year user of LSD) could not have first used hallucinogens within the past year, regardless of the hallucinogens recency value. The bound created using the child recency and frequency was calculated in exactly the same way as for the parent recency and frequency information. Data from the child drugs also were used to set a bound for the earliest possible date for date of first use under the following conditions: (1) respondents did not use any hallucinogens in their lifetime except for LSD, PCP, or Ecstasy; and (2) for each child drug used, respondents provided a valid year of first use. (If a valid month of first use was provided, this information also was used when calculating the bounds).

For hallucinogens child drugs (set 45), the imputation-revised hallucinogen DFU was used as an additional bound when setting the earliest possible date for the date of first use. This was done to ensure that the child date of first use was never earlier than the parent date of first use. Additionally, an indicator of lifetime hallucinogen drug use other than Ecstasy, PCP, and LSD was created. The algorithm for handling set 45 is detailed below:

1. The earliest possible date, latest possible date, and the final date of first use for each child drug for which the respondent was a lifetime user were assigned.
2. For respondents who were lifetime nonusers of other hallucinogens, it was determined which, if any, child drug could have had the same date of first use as hallucinogens. Specifically, it was determined whether the date of first use for hallucinogens was between the earliest possible date and latest possible date for each child drug. If none of the child drugs were eligible to receive the hallucinogens date of first use, nothing was done. Otherwise, one of the eligible child drugs was chosen at random, and its date of first use was overwritten with the hallucinogens date.

6.2.5.6.1.1 Final Variables

The final imputation-revised date-of-first-misuse variables were identified with the prefix IR, followed by a substance abbreviation and ending with an indicator of the measure (MFU = month of first use, YFU = year of first use, DFU = day of first use). Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were the standard levels used for most imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user). Because survey respondents are not asked for the specific day on which they first used the substance of interest, all respondents in the domain receive IIXXDFU = 3. Also, as indicated above, only recent initiates are asked for the year and month of first substance use. Subsequently, these questions have high rates of nonresponse because of the skip logic

embedded in the questionnaire, as all other individuals in the domain require imputation for their MFU and YFU.

6.2.5.6.2 *Age-at-First-Misuse and Date-of-First-Misuse Assignments for Prescription Drugs*

For age at first misuse and date of first misuse for prescription drugs, only past year initiates (i.e., respondents who misused the drug of interest for the first time in 12 months prior to the date of the interview) are included in the domain. Respondents who were not past year initiates receive an appropriate skip code. As a result, the age at first misuse (AFM) is limited to either the respondent's age or to the respondent's age – 1. Respondents were asked their AFM for each specific prescription drug they misused in the past 12 months. Recent initiates were asked for their month of first misuse (MFM) and year of first misuse (YFM). The day of first misuse (DFM) was not collected in the questionnaire and was missing for all respondents.

For each specific prescription drug within a given category, candidate upper and lower bounds on the date of first misuse (i.e., the earliest possible date of first misuse and the latest possible date of first misuse) were determined. Additionally, at the category level (i.e., overall pain relievers), additional candidate upper bounds are determined based on the interview date, recency of misuse, and the 30-day frequency of misuse. The overall category-level candidate lower bound is the minimum of all candidate lower bounds, and the final category-level upper bound is the minimum of all candidate upper bounds. For example, suppose a respondent was a past year initiate of two specific pain relievers with the following bounds: a lower bound of January 1 and an upper bound of January 31 for the first drug and a lower bound of March 1 and an upper bound of March 31 for the second drug. The pain reliever lower bound would be January 1 and the upper bound would be January 31. Note that if the maximum of the lower bound was taken, a negative interval would have been produced.

The final category-level lower bound is the maximum of the overall category-level candidate lower bound and 1 year before the interview date.¹³⁷ In the absence of any other available information to restrict the possible date of first misuse, the lower bound is set to 1 year before the interview date. Once the earliest and latest possible dates of first misuse for the category were determined, a day was randomly selected from this interval. The imputation-revised age, month, day, and year were extracted from this date of first misuse.

The process described above assumes equal probability for all days in the final interval; however, not all days in the interval should always have an equal probability. For example, suppose a respondent was interviewed on July 31, 2023, and indicated misusing two pain relievers in the past year, OxyContin® and Vicodin. The respondent indicated first misusing OxyContin® in March 2023 but left the year and month of first use for Vicodin as missing. The final category-level interval for possible days of initiation for pain relievers would be August 1, 2022, through March 31, 2023. Ignoring information provided by recency of misuse and 30-day frequency for simplicity, the respondent could have initiated Vicodin any day between August 1,

¹³⁷ Sometimes the candidate bound of 1 year before the interview date can cause a negative interval and must be modified. There were no instances for 2023.

2022, and July 31, 2023. Likewise, the respondent could have initiated OxyContin® any day between March 1, 2023, and March 31, 2023. Thus, the days occurring in March should actually have a higher probability of being the overall date of initiation for pain relievers than days that are not in March. Due to the complication in calculating these probabilities as the number of specific drugs misused increases and the possibility of incomplete information at the specific drug level, the decision was made to assume equal probability for all days in the final category-level interval.

6.2.5.6.2.1 Defining Upper and Lower Bounds

The candidate upper and lower bounds are defined as follows. Depending on the information provided by the past year initiate, some candidate bounds will be missing.

Candidate lower bounds =

- January 1 of YFM, if YFM provided
- First day of the month indicated by MFM/YFM, if MFM and YFM provided
- AFMth birthday, if AFM provided

Candidate upper bounds =

- December 31 of YFM, if YFM provided
- Last day of the month indicated by MFM/YFM, if MFM and YFM provided
- 1 day before (AFM + 1)th birthday, if AFM provided
- Interview date
- Interview date – category level 30-day frequency + 1, if category level recency is past month
- 30 days before the interview date, if category level recency is past year not past month

If a past year initiate provided no valid information, then the respondent could have initiated on any day in the past year, giving a lower bound of interview date – 1 year and an upper bound of the interview date.

If a past year initiate was interviewed on the last day of the month and the respondent reported first misuse of a specific drug in the same month as the interview in the previous calendar year, then the lower bound was set to the upper bound. That is, the date of first misuse was imputed to the last day of the month in the previous calendar year.

6.2.5.6.2.2 Final Variables

As with all other imputation-revised variables, the final imputation-revised age-at-first-misuse and date-of-first-misuse variables were identified with the prefix IR, followed by an eight-letter identifier, where a three-letter code identified the drug and the final five letters identified the measure (NMAge = age at first misuse, NMMFU = month of first misuse, NMYFU = year of first misuse, NMDFU = day of first misuse). Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, past year initiation imputed, 4 = statistically imputed, past year initiation known, and 9 = legitimate skip (not a

past year initiate). Because survey respondents are not asked for the specific day on which they first misused the drug of interest, all respondents in the domain receive IIXXXNMDFU = 3 or 4.

6.5.3 Recodes for Additional Analyses

From the final imputation-revised variables, some additional recodes were created to use in additional analyses. For recency-of-use variables, this included three dichotomous indicator variables to indicate lifetime (mis)use (XXX(NM)FLAG), past year (mis)use (XXX(NM)YR), and past month (mis)use (XXX(NM)MON). For the prescription drugs, two dichotomous indicator variables to indicate lifetime any use (XXXANYFLAG) or past year any use (XXXANYYR) were also created. Several other prevalence recodes, which covered the same three measures, were created to incorporate information from several different substances. The creation of these variables was also straightforward. If the respondent was a lifetime user of any of the substances, then the FLAG variable was set to 1; otherwise, it was set to 0. The YR and MON variables were processed in a similar manner. [Tables 6.18](#) and [6.19](#) contain information on all dichotomous indicator recodes created.

Table 6.18 Prevalence Recodes

Substance	Variable Names
Cigarettes	CIGFLAG, CIGYR, CIGMON
Smokeless Tobacco	SMKLSSFLAG, SMKLSSYR, SMKLSSMON
Cigars	CGRFLAG, CGRYR, CGRMON
Pipes	PIPFLAG, PIPMON
Alcohol	ALCFLAG, ALCYR, ALCMON
Marijuana	MRJFLAG, MRJYR, MRJMON
Cocaine	COCFLAG, COCYR, COCMON
Crack	CRKFLAG, CRKYR, CRKMON
Hallucinogens	HALLUCFLAG, HALLUCYR, HALLUCMON
LSD	LSDFLAG, LSDYR, LSDMON
PCP	PCPFLAG, PCPYR, PCPMON
Ecstasy	ECSTMOFLAG, ECSTMOYR, ECSTMOMON
Tryptamine Hallucinogens DMT, AMT, or "Foxy"	DAMTFXFLAG, DAMTFXYR, DAMTFXMON
<i>Salvia divinorum</i>	SALVIAFLAG, SALVIAYR, SALVIAMON
Ketamine	KETMINFLAG, KETMINYR, KETMINMON
Inhalants	INHALFLAG, INHALYR, INHALMON
Methamphetamine	METHAMFLAG, METHAMYR, METHAMMON
Pain Relievers	PNRANYFLAG, PNRANYR, PNRNMFLAG, PNRNMYR, PNRNMMON
Sedatives	SEDANYFLAG, SEDANYR, SEDNMFLAG, SEDNMYR, SEDNMMON
Stimulants	STMANYFLAG, STMANYR, STMNMFLAG, STMNMYR, STMNMMON
Tranquilizers	TRQANYFLAG, TRQANYR, TRQNMFLAG, TRQNMYR, TRQNMMON
Nicotine Vaping	NICVAPFLAG, NICVAPYR, NICVAPMON
CBD or Hemp	CBDHMPFLAG, CBDHMPYR, CBDHMPMON

Table 6.19 Prevalence Recodes Incorporating More Than One Recency Variable

General Substance Category	Variable Names	Source Recency Variables
Tobacco	TOBFLAG, TOBYR, TOBMON	Cigarettes, smokeless tobacco, cigars, pipes
Any Tobacco or Nicotine Vaping	TOBVNICFLAG, TOBVNICMYR, TOBVNICMON	Cigarettes, smokeless tobacco, cigars, pipes, nicotine vaping
Psychotherapeutics, Any Use	PSYANYFLAG, PSYANYR	Any use of pain relievers, tranquilizers, stimulants, sedatives
Psychotherapeutics, Misuse	PSYCHFLAG, PSYCHYR, PSYCHMON	Misuse of pain relievers, tranquilizers, stimulants, sedatives
Illicit Drugs Other than Marijuana	ILLEMFLAG, ILLEMYR, ILLEMMON	Psychotherapeutics (misuse), plus inhalants, hallucinogens, methamphetamine, cocaine, and heroin
Illicit Drugs, but Only Marijuana	MJONLYFLAG, MJONLYYR, MJONLYMON	Same as MRJFLAG, MRJYR, and MRJMON, except set to 0 if the corresponding ILLEM variable is equal to 1
Illicit Drugs	ILLFLAG, ILLYR, ILLMON	Illicit drugs other than marijuana, plus marijuana

For 30-day binge alcohol frequency of use, two recodes were created. The first, BNGDRKMON, is an indicator of any binge alcohol use in the past 30 days. The other, HVDYDRKMON, is an indicator of heavy alcohol use in the past 30 days. Heavy alcohol use is defined as drinking five or more drinks on the same occasion for males or four or more drinks on the same occasion for females on each of 5 or more days in the past 30 days.

For marijuana ways of use, the 16 imputed variables had values of 1 = yes, 2 = no, and 9 = did not use. For each of these variables, a corresponding dichotomous indicator was created that had values of 1 = yes and 0 = no. These variables were named MJxxxMON and MJxxxYR, where xxx was a 3-letter abbreviation for the way of use: SMK, VAP, DAB, EAT, MTH, PLL, SKN, and OTH.

7. Editing and Imputation for Special Drugs and Substance Use Disorder Variables

7.1 Introduction

This chapter discusses editing and imputation procedures for the special drugs section and substance use disorder (SUD) variables. The special drugs section, which was self-administered for in-person interviews, is included in this chapter because the variables for cocaine, heroin, and methamphetamine are used in editing the SUD variables for the corresponding drugs.¹³⁸

Section 7.2 describes the editing procedures for the variables in the special drugs section and, where applicable, the statistical imputation procedures used to replace missing data. Section 7.3 describes the editing and imputation procedures for the SUD variables and the nicotine (cigarette) dependence variables.

7.2 Special Drugs

The special drugs section asked about the use of GHB (gamma hydroxybutyrate) and the use of over-the-counter (OTC) cold medicines to get high. The special drugs section also asked about modes of drug use, including smoking and sniffing of heroin; use of heroin, cocaine, methamphetamine, or other drugs with a needle; general needle use behaviors (e.g., needle sharing); and where respondents got the last needle that they used.

7.2.1 GHB and OTC Cold Medicines

Question SD01 asked whether respondents ever used nonprescription cough or cold medicines (also known as OTC medicines) to get high. Question SD03 asked whether respondents ever used GHB (also called "G," "Georgia Home Boy," "Grievous Bodily Harm," "Liquid G," or gamma hydroxybutyrate). If respondents reported that they ever used OTC cough or cold medicines to get high or that they ever used GHB, they were asked to report when they last engaged in that behavior (questions SD02 and SD04, respectively).

In addition, respondents who reported that they last used OTC cough or cold medicines to get high in the past 12 months were asked to specify what OTC drugs they used to get high in that period. Respondents could specify up to five drugs that they used to get high. Therefore, the edited variables COLDYR1 through COLDYR5 (corresponding to the "OTHER, Specify" items SD02A through SD02E) captured up to five medications that respondents reported using to get high.

The "OTHER, Specify" variables for OTC cough and cold medicines were coded using a decimal system both to preserve relationships between similar responses and to capture additional details that respondents provided. For example, there are multiple formulations of OTC cold medicines under the brand name of Robitussin® that vary in terms of the specific ingredients they contain, such as the OTC cough suppressant dextromethorphan (DXM). Therefore, if

¹³⁸ All sections of the web questionnaire were self-administered.

respondents specified that they used a Robitussin[®] product to get high in the past 12 months but the only detail they provided was that they used "Robitussin," that response was assigned a default code of 829.0 (Robitussin[®], no other information). However, if respondents specifically reported that they used Robitussin[®] DM to get high (a product that contains DXM), that response was coded as 829.1 and was documented as Robitussin[®] DM. Thus, the main code of 829 indicated use of a Robitussin[®] product to get high in the past 12 months, and the decimal value indicated whether respondents provided additional detail about their use of specific Robitussin[®] products to get high.

An important aspect of the processing of these drug variables consisted of assigning codes to variables that had been skipped because the respondent had not used a particular drug in the relevant time period and the questions did not apply. For example, if respondents reported in question SD03 (edited variable GHB) that they never used GHB, the edited recency-of-use variable GHBREC (corresponding to question SD04) was assigned a code of 91. Similarly, if respondents reported that they last used an OTC cough or cold medicine to get high more than 12 months ago (edited variable COLDREC), COLDYR1 through COLDYR5 were assigned a code of 9993.0 to indicate that the questions did not apply because respondents used OTC cough or cold medicines to get high but not in the past 12 months.

In addition, if respondents reported that they had ever used these drugs but they did not know or refused to report when they last used the drug, the edited recency-of-use variable (e.g., GHBREC) was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). If the most recent use of OTC cough or cold medicines (edited variable COLDREC, corresponding to question SD02) had been assigned a code of 9, COLDYR1 through COLDYR5 were left as blank (code of 9998.0) because it was unknown whether these respondents used OTC cough or cold medicines to get high in the past 12 months. The questions would have been applicable if respondents last used OTC cough and cold medicines to get high in the past 12 months.

Data from the "OTHER, Specify" needle use variables OTDGNDLA through OTDGNDLE from questions SD15SP to SD15D (Section 7.2.5.2) and the cough and cold medicine variables COLDYR1 through COLDYR5 were used in editing GHB and GHBREC. However, OTDGNDLA through OTDGNDLE were not used to edit COLDMEDS (corresponding to question SD01) and COLDREC (corresponding to SD02) because there could be too much room for interpretation regarding what might constitute a nonprescription cough or cold medicine. In contrast, the questions for the other additional drugs in the special drugs section pertained to specific drugs.

Thus, for example, if respondents had not reported lifetime use of GHB but specified using GHB with a needle in OTDGNDLA through OTDGNDLE, then the edited variable GHB (corresponding to question SD03) would be assigned a code of 3 (Yes LOGICALLY ASSIGNED). In turn, the recency-of-use variable GHBREC (corresponding to question SD04) would be assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). However, if respondents had not reported lifetime use of OTC cold medicines to get high but specified using a specific cold medicine with a needle in OTDGNDLA through OTDGNDLE, then no edits were made to COLDMEDS or COLDREC.

If respondents specified use in the past year of GHB as a nonprescription cough or cold medicine in COLDYR1 through COLDYR5, then the corresponding edited variable GHB was assigned a code of 3 (Yes LOGICALLY ASSIGNED), but GHBREC was assigned a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) because COLDYR1 through COLDYR5 pertain to use in the past 12 months. If respondents had answered question SD03 for GHB as “yes” and reported last using it more than 12 months ago but also specified using it in the past 12 months in COLDYR1 through COLDYR5, then GHBREC also was assigned a code of 8.

Because the questions corresponding to COLDMEDS and COLDREC asked specifically about use of OTC cough or cold medicines, respondents who reported using OTC cough or cold medicines in the past 12 months to get high also would be expected to report at least one OTC medicine in COLDYR1 through COLDYR5 or to have some response of “don’t know” or “refused” in the corresponding items (i.e., if respondents did not know the name of the drug or did not want to report it). Therefore, a variable called OTCFLAG was created to indicate when respondents’ reports of use of OTC cough or cold medicines in the past year were not substantiated by the “OTHER, Specify” data.¹³⁹ OTCFLAG had the values shown below when these situations occurred.

- OTCFLAG had a value of 1 when respondents specified only the use of drugs that require a prescription in the United States or drugs that are not available as OTC medicines in the United States, even if they may be available as OTC medicines elsewhere.
- OTCFLAG had a value of 2 when COLDYR1 through COLDYR5 had values only of 9985.0 (bad data) or 9998.0 (blank).

No further editing was done to the OTC variables when OTCFLAG indicated that respondents’ reports of use of OTC cough or cold medicines to get high in the past year were not substantiated in COLDYR1 through COLDYR5. However, OTCFLAG would alert analysts when this issue occurred.

7.2.2 General Edits of Cocaine, Heroin, and Methamphetamine Variables for Nonusers

An important aspect of the processing of variables in this section—pertaining to the use of cocaine, heroin, or methamphetamine with a needle or the smoking or sniffing of heroin—consisted of assigning codes of 91, 93, or 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, respondents who never used cocaine, heroin, or methamphetamine were not asked questions in the special drugs section that

¹³⁹ If respondents reported the use of drugs that may require a prescription in some locations in the United States but may be available without a prescription elsewhere in the United States (e.g., cough medicines containing low doses of codeine), the medication was assumed to have been obtained without a prescription.

pertained to these drugs. [Table B.11](#) in Appendix B describes specific edits that were implemented for these variables when the corresponding items were skipped.

7.2.3 Edits for Heroin Users Who Did Not Report Smoking, Sniffing, or Injection of Heroin

Respondents who reported in the earlier heroin section of the interview that they were lifetime heroin users but who answered “no” to all questions about smoking heroin (question SD07), sniffing heroin (question SD09), or using it with a needle (question SD11) were asked a follow-up question SDHEUSE to determine how these respondents administered the heroin they had reported using. SDHEUSE was an “enter all that apply” type of question that allowed respondents to report multiple ways that they used heroin. SDHEUSE included response options for smoking heroin (edited variable HEOTSMK), sniffing heroin (edited variable HEOTSNF), using heroin with a needle (edited variable HEOTNDL), or use of heroin “some other way” (edited variable HEOTSP). Respondents who reported using heroin only in “some other way” were asked to specify in question SDHEUSE2 what this “other” mode of heroin administration was. If respondents specified using heroin in a way that corresponded to one of the ways that they had been asked about, it was logically inferred that the response in the edited SDHEUSE question should have been chosen. For example, if respondents did not report in SDHEUSE that they smoked heroin but they specified smoking it in “some other way,” the edited variable HEOTSMK was assigned a code of 3 (Response entered LOGICALLY ASSIGNED).

In turn, HEOTSMK, HEOTSNF, and HEOTNDL were used to apply logically assigned values to the corresponding lifetime heroin variables HERSMOKE (SD07), HERSNIFF (SD09), and HERNEEDL (SD11), respectively. For example, if HEOTSMK indicated that the respondents had smoked heroin (HEOTSMK = 1 or 3), HERSMOKE was edited to indicate that the respondents were logically inferred to have smoked heroin at least once in their lifetime.

7.2.4 Additional Edits for Smoking or Sniffing of Heroin

As noted previously, respondents who reported that they had ever used heroin were asked in questions SD07 and SD09 to report whether they had ever smoked or sniffed heroin (i.e., inhaled heroin powder through their nose), respectively. Respondents were routed to follow-up recency questions for smoking or sniffing of heroin if they answered the corresponding lifetime questions as “yes” or if they reported using heroin in either of these ways in question SDHEUSE (Section 7.2.3). Information from the follow-up recency questions SD08, SD10, SDHEUSE3, and SDHEUSE4 were used in the creation of the heroin smoking and sniffing recency variables HRSMKREC and HRSNFREC. If respondents did not report lifetime use of heroin in these ways (e.g., SD07 for smoking heroin was answered as “no”) and smoking or sniffing of heroin was not reported in SDHEUSE, then HRSMKREC and HRSNFREC were assigned codes of 93 to indicate that the respondents had used heroin but had not smoked it or had not sniffed it, respectively.

Also, if respondents initially did not report smoking or sniffing heroin in SD07, SD09, or SDHEUSE, then they were skipped out of the follow-up recency questions for smoking and sniffing of heroin. Therefore, if respondents’ only indication of smoking or sniffing of heroin

came from the “OTHER, Specify” response associated with SDHEUSE, the corresponding variables HRSMKREC and HRSNFREC were set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED).

Consistent with the principles described in Section 2.4.1 for editing across sections, no editing was done to the heroin data in an earlier section of the interview if lifetime heroin users reported more recent smoking or sniffing of heroin (edited variables HRSMKREC or HRSNFREC, respectively) than they reported for when they last used heroin (edited variable HERREC). Consequently, HRSMKREC or HRSNFREC could indicate more recent use of heroin than was indicated in the edited heroin recency variable HERREC or in the imputed heroin recency variable IRHERRC.

[Table B.12](#) describes specific edits other than those pertaining to skip logic that were implemented in the special drugs section for smoking and sniffing of heroin. For example, lifetime users of heroin could report that they smoked heroin at least once but not indicate when they last smoked it. The general edit was to assign a nonspecific value to the edited recency variable (i.e., HRSMKREC) to indicate that the respondents smoked heroin at some point in their lifetime. In some special situations, however, it was possible to infer that respondents could not have smoked heroin in the past 12 months. In these situations, respondents reported last using heroin more than 12 months ago, and there were no responses for other heroin-related questions in the special drugs section to indicate that these respondents had used heroin in the past 12 months.

7.2.5 Injection Drug Use

Respondents who reported lifetime use of the relevant substance were asked whether they had ever used a needle to inject cocaine (question SD05), heroin (questions SD11 or SDHEUSE), or methamphetamine (SD13). Respondents were also asked in question SD15 whether they used a needle to inject other drugs (if they reported injecting cocaine, heroin, or methamphetamine) or any drug (otherwise). Respondents who reported lifetime injection of a drug were routed to follow-up recency questions (SD06 for cocaine, SD12 for heroin, SD14 for methamphetamine, and SD16 for other drugs). Information from these questions was used in the creation of the corresponding needle recency variables CONDLREC, HRNDLREC, METHNDLRC, and OTDGNLRC.

7.2.5.1 Injection of Cocaine, Heroin, or Methamphetamine

[Table B.12](#) describes specific edits other than those pertaining to skip logic that were implemented in the special drugs section for the use of drugs with a needle. For example, lifetime users of cocaine could report that they used cocaine with a needle at least once but not indicate when they last injected it. The general edit was to assign a nonspecific value to the edited recency variable (i.e., CONDLREC) to indicate that the respondents used cocaine with a needle at some point in their lifetime. In some special situations, however, it was possible to infer that respondents could not have injected cocaine in the past 12 months. In these situations, respondents reported last using cocaine more than 12 months ago in the cocaine

section, and there were no responses for other cocaine-related questions in the special drugs section to indicate that these respondents had used cocaine in the past 12 months.

Consistent with the principles described in Section 2.4.1 for editing across sections, no editing was done to the corresponding data in earlier sections of the interview if respondents reported more recent use of cocaine, heroin, or methamphetamine with a needle than they previously reported for when they last used these drugs. Consequently, CONDLREC, HRNDLREC, or METHNDLRC could indicate more recent use of these drugs than was indicated in the corresponding edited or imputed recency variables (e.g., HERREC and IRHERRC for heroin).

As noted previously, respondents also were asked in question SD15 (edited variable OTDGNEDL) whether they ever used a needle to inject “some other drug” (if respondents previously reported lifetime use of cocaine, heroin, or methamphetamine with a needle) or “any drug” (if respondents had not previously indicated use of any of these drugs with a needle). If question SD15 was answered as “yes” (OTDGNEDL = 1), respondents were asked to specify what (other) drug(s) they used with a needle. Respondents could specify up to five drugs that they had injected (edited variables OTDGNDLA through OTDGNLDE).

Consequently, it was possible for respondents to have reported in a previous substance use section that they never used cocaine, heroin, or methamphetamine but then to specify use of that drug with a needle in OTDGNDLA through OTDGNLDE. For example, respondents could indicate in the earlier heroin section that they never used heroin but then specify lifetime injection of heroin in OTDGNDLA through OTDGNLDE.

Consistent with the principles described in Section 2.4.1 for editing across sections, no editing was done to the corresponding data in earlier sections of the interview if respondents reported that they never used cocaine, heroin, or methamphetamine but they reported using any of these drugs with a needle in the special drugs section of the interview. However, these respondents were logically inferred in the relevant special drugs variables to be users of that particular drug at some point in their lifetime. Consequently, CONDLREC, HRNDLREC, or METHNDLRC could indicate use of cocaine, heroin, or methamphetamine, respectively, even if the corresponding edited or imputed variables from earlier sections of the interview indicated lifetime nonuse. Thus, if respondents reported in the heroin section that they never used heroin, for example, but then they specified heroin as “some other drug” that they used with a needle, then the edited lifetime heroin needle use variable HERNEEDL was assigned a code of 3 (Yes LOGICALLY ASSIGNED). The edited heroin recency variable HERREC retained a value of 91 to indicate that respondents previously reported that they never used heroin. If these respondents indicated in SD16 that they last used a needle to inject all other drugs more than 12 months ago (OTDGNLRC = 3), then it could be inferred that their most recent use of heroin with a needle occurred more than 12 months prior to the interview date (HRNDLREC = 13). Otherwise HRNDLREC was assigned a code of 9 (Used at some point in lifetime LOGICALLY ASSIGNED).

Respondents also could specify in OTDGNDLA through OTDGNLDE that they injected only cocaine, only heroin, or only methamphetamine in their lifetime and then report more recent use of this “other” drug with a needle in SD16 (edited variable OTDGNLRC) than they had

reported in the corresponding needle recency questions for cocaine, heroin, or methamphetamine. When this pattern occurred, CONDLREC, HRNLDREC, or METHNDLRC were assigned the value of most recent injection of other drugs from OTDGNDLRC, plus a value of 10 (i.e., “bump” by 10). For example, if respondents reported last injecting cocaine more than 12 months ago (SD06 = 3), specified the lifetime injection of only cocaine, and reported last injecting other drugs more than 30 days ago but within the past 12 months (OTDGNDLRC = 2), then CONDLREC was assigned a code of 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).

This edit also was implemented if respondents reported lifetime injection of cocaine, heroin, or methamphetamine but the most recent time when respondents injected the drug was unknown. For example, if CONDLREC = 9 (At some point in the lifetime) and OTDGNDLRC = 3 (More than 12 months ago) when injection of cocaine was the only “other” drug specified, then CONDLREC was assigned a code of 13. In this situation, it could be inferred that the respondent’s most recent use of cocaine with a needle occurred more than 12 months prior to the interview date.

This edit was not implemented if respondents specified that they injected other drugs besides cocaine, heroin, or methamphetamine in their lifetime¹⁴⁰ or they injected more than one of these specific drugs in their lifetime. In these situations, it was not possible to infer that more recent injection of some other drug applied specifically to cocaine, heroin, or methamphetamine.

7.2.5.2 Injection of Other Drugs

As noted previously, respondents could report in question SD15 that they used a needle to inject other drugs, or that they used a needle to inject any drug, if they had not previously reported using a needle to inject cocaine, heroin, or methamphetamine. Respondents who answered SD15 as “yes” were asked to specify the names of up to five other drugs that they injected in their lifetime.

Respondents could report that they injected other drugs in their lifetime, even though they previously reported that they never used marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine, and that they had never misused prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, or sedatives). However, the data from OTDGNDLA through OTDGNDLE could be used to identify respondents who corroborated their report of lifetime injection drug use (e.g., if injection of anabolic steroids was reported) despite having previously reported nonuse (or no misuse) of all drugs covered in the previous substance use sections. Similarly, from “OTHER, Specify” data on other drugs that respondents had injected, it was possible to identify those whose needle use was probably limited to legitimate, medical uses (e.g., injection of antibiotics). Therefore, it was logically inferred that respondents had never used a needle to inject other drugs if (1) they were lifetime nonusers of all drugs covered

¹⁴⁰ This situation also applied to respondents who reported only cocaine, heroin, or methamphetamine as specific other drugs they injected but they also had responses of “don’t know” or “refused” in their “OTHER, Specify” data. The responses of “don’t know” or “refused” were assumed to be potential indications that respondents injected other drugs besides cocaine, heroin, or methamphetamine.

in the sections for the substances that were described previously; (2) they indicated that they never engaged in behaviors that would be indicative of the misuse of drugs with a needle, such as needle sharing, use of bleach to clean needles, or reusing of needles; and (3) all of the “other” drugs they reported using with a needle typically were not drugs with abuse potential. In this situation, the variable OTDGNEDL corresponding to question SD15 was set to a value of 4 (No LOGICALLY ASSIGNED). Edits to the general needle use variables when OTDGNEDL had been assigned a code of 4 are discussed in Section 7.2.5.4.

Just as data from question SD16 for the most recent injection of other drugs could be used to infer more recent injection in the edited variables CONDLREC, HRNDLREC, and METHNDLRC, data from CONDLREC, HRNDLREC, and METHNDLRC could be used to infer more recent injection in OTDGNLRC. Similar to the edits described in Section 7.2.5.1, data from CONDLREC, HRNDLREC, and METHNDLRC were used to edit OTDGNLRC when respondents had specified the lifetime injection of only cocaine, only heroin, or only methamphetamine in OTDGNDLA through OTDGNLDE. For example, if respondents reported the lifetime injection of only cocaine, reported last injecting cocaine as “some other drug” more than 12 months ago (SD16 = 3), but they reported using a needle to inject cocaine more than 30 days ago but within the past 12 months (CONDLREC = 2), then OTDGNLRC was assigned a code of 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).

7.2.5.3 Recoded Needle Use Variables

The recoded needle use variable CHMNDLREC (where “CHM” refers to cocaine, heroin, or methamphetamine) was created from variables pertaining to the use of cocaine, heroin, or methamphetamine with a needle. Likewise, ANYNDLREC (any needle use) was created from variables pertaining to the use of cocaine, heroin, methamphetamine, or other drug with a needle.

Specifically, the needle recency variable CHMNDLREC was created from the variables CONDLREC (most recent use of cocaine with a needle), HRNDLREC (most recent use of heroin with a needle), and METHNDLRC (most recent use of methamphetamine with a needle). The logic for creating CHMNDLREC is documented below.

- If respondents never used a needle to inject cocaine, heroin, or methamphetamine (including situations in which respondents never used any of these drugs), CHMNDLREC indicated that the respondents had never used cocaine, heroin, or methamphetamine with a needle.
- If respondents reported using one or more of these drugs with a needle, the general principle in assigning a value to CHMNDLREC was to identify the most recent use reported by the respondents. In particular, if respondents reported using cocaine, heroin, or methamphetamine with a needle in the past 30 days, it could be determined unambiguously that the respondents were past month needle users.
- In other situations, however, if one or more of the cocaine, heroin, or methamphetamine needle recency variables indicated nonspecific use at some point in the respondents’ lifetime, CHMNDLREC was assigned a value to indicate nonspecific past year or lifetime use.

- For example, if respondents reported last using cocaine with a needle more than 30 days ago but within the past 12 months, yet all that was known was that they used heroin with a needle at some point in their lifetime, it still could be inferred that they had used some drug with a needle in the past 12 months. The nonspecific value for past year use (i.e., a code of 8) was assigned because the respondents could have been more recent users of heroin with a needle.
- Similarly, if respondents indicated use of one of these drugs with a needle more than 12 months ago and they did not know or refused to indicate when they last used one of the other drugs with a needle, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to CHMNDLREC because the respondents may have been more recent needle users than what they had reported elsewhere.

The recoded needle use variable ANYNDLREC (most recent use of any drug with a needle) also was created to take into account data from CONDLREC, HRNDLREC, METHNDLRC, and OTDGNDLRC. The same general principles that were described above for the creation of CHMNDLREC were used to create ANYNDLREC. For example, if CONDLREC, HRNDLREC, METHNDLRC, or OTDGNDLRC indicated use of drugs with a needle, the general principle in assigning a value to ANYNDLREC was to identify the most recent use reported by the respondents. In particular, if CONDLREC, HRNDLREC, METHNDLRC, or OTDGNDLRC indicated use of a drug with a needle in the past 30 days, then ANYNDLREC was assigned a code of 1 (Within the past 30 days). In this situation, it could be determined unambiguously that the respondents were past month needle users.

Consistent with the general principle of not editing across sections (Section 2.4.1), ANYNDLREC used data from the needle recency variables in the special drugs section of the interview. However, ANYNDLREC did not take into account data from respondents who reported in the prescription stimulants section of the interview that they injected prescription stimulants in the past 12 months (STMNDLYR = 1) or respondents who reported in the emerging issues section of the interview that they injected illegally made fentanyl in the past 12 months (IMFREC = 1 or 2). Unlike the data in the special drugs section for the most recent use of needles to inject drugs, the needle use data from the prescription stimulants section covered only the past 12 months and past 30 days but did not give respondents an opportunity to report last injecting prescription stimulants more than 12 months ago. If respondents specified the injection of prescription stimulants in OTDGNDLA through OTDGNDLE, however, then the most recent injection of other drugs from OTDGNDLRC (used in creating ANYNDLREC) could apply to prescription stimulants or illegally made fentanyl.

7.2.5.4 General Needle Use

Questions SD17 through SD20 asked lifetime needle users about general needle use behaviors, including reuse of a needle that the respondent had used before (SD17; edited variable GNNDREUS), use of a needle that the respondent knew or suspected had been previously used by someone else (SD18; edited variable GNNDLSH1), use of bleach to clean the needle the last time that the respondent used a needle to inject drugs (SD19; edited variable GNNDCLEN), and use of the needle by someone else after the respondent last used a needle (SD20; edited

variable GNNDLSH2). In addition, respondents were asked in question SD21 (and in the corresponding "OTHER, Specify" variable, if applicable) to report how they got the last needle that they used for injecting drugs (edited variable GNNDGET).

An important aspect of the processing of variables for general needle use behaviors or the source of the last needle that respondents used consisted of assigning a legitimate skip code of 99 if respondents had never used a needle to inject drugs (Section 2.4.2). In particular, if respondents did not report that they used a needle to inject prescription stimulants, cocaine, heroin, or methamphetamine, then question SD15 asked whether respondents ever used a needle to inject any drug. Even if respondents answered the needle use questions for prescription stimulants, cocaine, heroin, or methamphetamine as "don't know" or "refused," a response of "no" in SD15 could be interpreted to mean that the respondent had never used a needle to inject any drug. Consequently, legitimate skip codes were assigned to GNNDREUS through GNNDGET in this situation when SD15 was answered as "no" (OTDGNEDL = 2).

When OTDGNEDL had been set to a value of 4 as described in Section 7.2.5.2, any data in the general needle use variables GNNDREUS through GNNDGET were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned) to indicate that respondents logically should have skipped these items because they appeared to be reporting about legitimate use of drugs with a needle. However, this edit was not implemented if respondents answered any of the general needle use questions as "yes" because affirmative answers to these questions would indicate the misuse of drugs with a needle.

7.2.6 Imputation of Variables for GHB and OTC Cold Medicines

Variables for GHB and OTC cold medicines were imputed using the modified predictive mean neighborhood (modPMN) methodology as described in Chapter 3. Section 7.2.6.1 describes how these variables were segregated into imputation sets, Section 7.2.6.2 describes how the variables were ordered within the imputation sets, and Section 7.2.6.3 provides details on the modPMN steps and the sequencing of steps.

For GHB, the base variable used for imputation was GHBREC. For OTC cold medicines, the base variable for imputation was created by combining the edited variables COLDMEDS and OTCFLAG. If OTCFLAG was equal to 1 or 2, then the base variable was set to missing and subsequently imputed. Otherwise, the base variable for imputation was set to be equal to the value for COLDMEDS. See Section 7.2.1 for more information on the creation of GHBREC, COLDMEDS, and OTCFLAG.

7.2.6.1 Imputation Sets

The special drugs variables imputed using the modPMN methodology were grouped into a single imputation set as outlined in [Table 7.1](#).

Table 7.1 Substances and Substance Use Measures

Substance	Lifetime Usage/Recency of Use
GHB	Set 1
Over-the-Counter Cold Medicines	

7.2.6.2 Imputation Order within Imputation Sets

As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rates. The ordering for 2023 is shown in [Table 7.2](#).

Table 7.2 Sequencing of Substances within Imputation Sets, 2023

Imputation Set	Sequence
Lifetime Usage/Recency of Use: GHB and Over-the-Counter Cold Medicines	GHB, Over-the-Counter Cold Medicines

7.2.6.3 Imputation Process

The imputation procedures used to impute GHB and OTC medicines are described in detail in Section 6.5.2.3. Deviations from the process described in Section 6.5.2.3 are described below.

7.2.6.3.1 Response Propensity Step

The overall process for the response propensity step was similar to the process described in Section 6.5.2.3.1. However, the number of models fit and the inputs to the response propensity step can vary across imputation sets. The total number of response propensity models fit for the special drugs variables set is shown in [Table 7.3](#). The four key inputs to the response propensity adjustment step are described below.

- **Analysis Weight:** The preliminary analysis weight was used.
- **Domain:** The domain for this set is detailed in [Table 7.3](#).
- **Item Response Indicator:** If a valid response was provided for the given measure for the substance, the individual was deemed an item respondent.
- **Covariates:** Appendix D provides information on the covariates used in each of the response propensity models.

Table 7.3 Summary of Response Propensity Models

Set	Number of Response Propensity Models	Domain
1: Lifetime Usage/Recency of Use; GHB, Over-the-Counter Cold Medicines	2	All respondents

7.2.6.3.2 Prediction Modeling Step

Section 6.5.2.3.2 details the SUDAAN® method used to model multilevel categorical variables like those in the special drugs set. The levels for outcome variables for the special drugs set are listed in [Table 7.4](#).

Table 7.4 Outcomes of Prediction Modeling

Set	Outcome Variable Levels
1: Lifetime Usage/Recency of Use; GHB, Over-the-Counter Cold Medicines	1: Past Month Use 2: Past Year, Not Past Month Use 3: Lifetime, Not Past Year Use 4: Never Used

7.2.6.3.3 Collocated Stochastic Imputation Step

There are no deviations from the collocated stochastic imputation procedure described in Section 6.5.2.3.3. In the circumstance of partial information, the predictive mean vector is adjusted to prevent inconsistent imputation. For example, if a respondent reported lifetime usage of GHB but failed to answer the recency-of-use question, the predicted probability associated with “never used” is set to 0. The remaining probabilities corresponding to past month use, past year not past month use, and lifetime not past year use are then rescaled to sum to 1 prior to imputation. For more information on these restrictions, see Appendix E.

7.2.6.3.4 Final Variables

No final imputation-revised variables indicating lifetime usage alone were created because this information was recoded in the final imputation-revised recency-of-use variables. Imputation indicators also were not created.

The final imputation-revised variables for the special drugs set were identified with the IR prefix, followed by a drug abbreviation and ending with REC to indicate recency of use (i.e., IRGHBREC, IRNPCOLDREC). Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

7.2.6.4 Recodes for Additional Analyses

From the final imputation-revised variables, some additional recodes were created to use in additional analyses. This included three dichotomous indicator variables to indicate lifetime use (XXXFLAG), past year use (XXXYR), and past month use (XXXMON). [Table 7.5](#) contains information on the dichotomous indicator recodes created.

Table 7.5 Prevalence Recodes

Drug	Variable Names
GHB	GHBFLAG, GHBYR, GHBMON
Over-the-Counter Cold Medicines	NPCOLDFLAG, NPCOLDYR, NPCOLDMON

7.3 Substance Use Disorder

Beginning with the 2021 National Survey on Drug Use and Health (NSDUH), the SUD questions are consistent with criteria from the *Diagnostic and Statistical Manual of Mental Disorders*, 5th edition (DSM-5; American Psychiatric Association, 2013). Respondents were asked about symptoms of SUD in the past 12 months that were associated with the use of alcohol, marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, methamphetamine, or any use of prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives) in that period. Lifetime users of cocaine (including crack), heroin, and methamphetamine were also asked questions about SUD for the relevant drug if they reported the use of these drugs in the past 12 months in the special drugs section of the interview (Sections 7.2.4 and 7.2.5). Respondents were asked the SUD questions for alcohol and marijuana only if they reported use of these substances on 6 or more days in the past year.

In addition, the SUD section of questionnaire included items to assess for dependence on cigarettes if respondents reported use of cigarettes in the past 30 days. Details on how respondents were classified as having nicotine (cigarette) dependence or SUDs for alcohol or drugs are discussed in Section 3.4.3 in the *2023 National Survey on Drug Use and Health (NSDUH): Methodological Summary and Definitions* (Center for Behavioral Health Statistics and Quality, 2024a).

7.3.1 Editing of Substance Use Disorder and Cigarette Dependence Variables

Section 7.3.1.1 describes the editing procedures in place for SUD variables for respondents who were nonusers in the period of interest. Section 7.3.1.2 describes the edits applied to the questionnaire items that were used to develop the composite score for determining cigarette dependence. Section 7.3.1.3 describes the assignment of a legitimate skip code for alcohol through sedative use disorder variables. Sections 7.3.1.4 and 7.3.1.5 describe the creation of DSM-5 SUD indicators and SUD severity measures.

7.3.1.1 Editing of Substance Use Disorder Variables for Nonusers in the Period of Interest

An important aspect of editing the SUD variables consisted of assigning codes of 91 or 93 (Section 2.4.2) to variables that had been skipped because respondents never used the substance or they used it but not in the period of interest (i.e., more than 30 days ago for cigarettes or

more than 12 months ago for other substances).¹⁴¹ For example, if respondents never used a substance, then codes of 91 were assigned to the edited SUD variables.

If recency-of-use variables for the use of psychotherapeutic drugs were assigned a code of 83 (i.e., Did not use in the past 12 months but lifetime use was unknown), then any SUD data for that psychotherapeutic drug were assigned a code of 83. For cocaine, heroin, and methamphetamine, respondents' answers to SUD questions were retained if respondents were routed into the respective set of SUD questions because they reported past year use in the special drugs section (Sections 7.2.4 and 7.2.5).

For alcohol and marijuana, respondents who had used these substances in the past 12 months also were skipped out of the corresponding SUD questions if they were only infrequent users of these two substances in the past 12 months (i.e., use on fewer than 6 days). Therefore, for SUD variables for alcohol and marijuana among respondents, the final edited 12-month frequency variables (ALCYRTOT and MJYRTOT, respectively) also were used in assigning codes of 93 or 83 to the SUD variables pertaining to these substances. For example, if the edited variable ALCYRTOT indicated that respondents had used alcohol in the past 12 months but on fewer than 6 days in that period, then the edited SUD variables for alcohol were assigned a code of 93 if they had been skipped. If respondents answered one or more SUD questions for alcohol, but the final value for ALCYRTOT indicated that they had used alcohol on fewer than 6 days in the past 12 months, then the previous answers in the SUD questions were overwritten with a code of 83. Similar edits were done for marijuana if MJYRTOT indicated that respondents used marijuana on fewer than 6 days in the past 12 months.

7.3.1.2 Editing of Cigarette Dependence Variables

NSDUH included questions designed to assess cigarette dependence (also known as "nicotine dependence") based on two methods. The first method for determining cigarette dependence involved calculating a continuous score from the Nicotine Dependence Syndrome Scale (NDSS) (Shiffman et al., 1995; Shiffman et al., 2003). The score was calculated from the 17 NDSS questions shown in [Table 7.6](#) that were asked of respondents who used cigarettes in the past 30 days. For each of these items, respondents who had used cigarettes in the past 30 days were asked to choose an answer from the following scale: 1 = Not at all true; 2 = Somewhat true; 3 = Moderately true; 4 = Very true; or 5 = Extremely true. The second method for determining cigarette dependence was based on the Fagerstrom Test for Nicotine Dependence (FTND). Because most nicotine is cleared from the bloodstream overnight, smokers typically wake in nicotine deprivation, and rapid movement to smoke is considered a sign of dependence. Respondents who reported smoking cigarettes in the past month were classified

¹⁴¹ For cocaine, heroin, and methamphetamine, respondents were not asked the corresponding SUD questions if there was no indication of use in the past 12 months either in the relevant earlier substance use section (or sections, in the situation of cocaine and crack) or in respondents' answers from the special drugs section. As noted in the main text, however, respondents who did not indicate past year use of cocaine, heroin, or methamphetamine in the corresponding earlier substance use sections but indicated past year use in the special drugs section *were* routed by the CAI instrument into the relevant SUD questions.

as nicotine dependent based on the FTND if they reported that they first smoked a cigarette within 30 minutes of waking up on the days that they smoked. For details on how the estimate for cigarette dependence was calculated based on the NDSS and FTND, see the nicotine dependence and substance use disorder appendix of the *2023 National Survey on Drug Use and Health Public Use File Codebook* (Center for Behavioral Health Statistics and Quality, 2024b).

As for other variables in the substance use disorder section, editing of the cigarette dependence variables also involved assignment of a legitimate skip code of 99 according to the skip logic within this section. For example, question DRCGE06A (not shown in [Table 7.6](#)) asked whether respondents who reported cigarette use in the past 30 days had any friends who did *not* smoke cigarettes. If DRCGE06A was answered as “no,” question DRCGE06B was skipped (i.e., choosing at times not to be around friends who do not smoke because they do not like it when the respondent smokes). The edited variable CIGFNLKE corresponding to DRCGE06B was assigned a code of 99 if DRCGE06A was answered as “no.”¹⁴²

Table 7.6 Mapping of Cigarette Dependence Questions to Edited Variables

Question	Edited Variable	Cigarette Dependence Scale	Question Text
DRCGE01	CIGIRTL	NDSS	After not smoking for a while, you need to smoke in order to feel less restless and irritable.
DRCGE02	CIGCRAVE	NDSS	When you don’t smoke for a few hours, you start to crave cigarettes.
DRCGE03	CIGCRAGP	NDSS	You sometimes have strong cravings for a cigarette where it feels like you’re in the grip of a force you can’t control.
DRCGE04	CIGINCTL	NDSS	You feel a sense of control over your smoking – that is, you can “take it or leave it” at any time.
DRCGE05	CIGAVOID	NDSS	You tend to avoid places that don’t allow smoking, even if you would otherwise enjoy them.
DRCGE07	CIGPLANE	NDSS	Even if you’re traveling a long distance, you’d rather not travel by airplane because you wouldn’t be allowed to smoke.
DRCGE08	CIGRNOUT	NDSS	You sometimes worry that you will run out of cigarettes.
DRCGE09	CIGREGDY	NDSS	You smoke cigarettes fairly regularly throughout the day.
DRCGE10	CIGREGWK	NDSS	You smoke about the same amount on weekends as on weekdays.
DRCGE11	CIGREGNM	NDSS	You smoke just about the same number of cigarettes from day to day.
DRCGE12	CIGNMCHG	NDSS	It’s hard to say how many cigarettes you smoke per day because the number often changes.
DRCGE13	CIGSVLHR	NDSS	It’s normal for you to smoke several cigarettes in an hour, then not have another one until hours later.
DRCGE14	CIGINFLU	NDSS	The number of cigarettes you smoke per day is often influenced by other things – how you’re feeling or what you’re doing, for example.

¹⁴² Consistent with the principles in Section 2.4.2, CIGFNLKE retained missing values if DRCGE06B was skipped because respondents answered DRCGE06a as “don’t know” or “refused.”

Table 7.6 Mapping of Cigarette Dependence Questions to Edited Variables (continued)

Question	Edited Variable	Cigarette Dependence Scale	Question Text
DRCGE15	CIGNOINF	NDSS	Your smoking is not affected much by other things. For example, you smoke about the same amount whether you're relaxing or working, happy or sad, alone or with others.
DRCGE16	CIGINCRS	NDSS	Since you started smoking, the amount you smoke has increased.
DRCGE17	CIGSATIS	NDSS	Compared to when you first started smoking, you need to smoke a lot more now in order to be satisfied.
DRCGE18	CIGLOTMR	NDSS	Compared to when you first started smoking, you can smoke much, much more now before you start to feel anything.
DRCGE19A, DRCGE19B	CIGWAKE	FTND	On the days that you smoke, how soon after you wake do you have your first cigarette?

7.3.1.3 Editing of Substance Use Disorder Variables for Alcohol through Sedatives

In addition to the assignment of codes of 91 or 93 for lifetime nonusers or respondents who last used a substance more than 12 months prior to the interview, an important aspect of the processing of the SUD variables for alcohol through sedatives involved assignment of a legitimate skip code of 99 (Section 2.4.2). This code was assigned when respondents qualified for being asked SUD questions about a given substance but they legitimately skipped out of one or more questions about that substance.

For example, the symptom of tolerance to the effects of alcohol was measured through two related questions, DPALMORE ("Do you need to drink a lot more alcohol than you used to in order to get the feeling you want?") and DPALLESS ("Does drinking the same amount of alcohol have much less effect on you than it used to?"). An affirmative answer to either question would indicate tolerance. Thus, if respondents had already answered DPALMORE as "yes," there was no need to ask DPALLESS. If the edited variable corresponding to question DPALMORE (UDALNEEDMOR) was coded as 1 (i.e., "yes"), the edited variable corresponding to question DPALLESS (UDALNEEDMOR) was assigned a legitimate skip code.

7.3.1.4 Creation of Cigarette Dependence Indicators

This section describes how the edited cigarette dependence indicators were created. For each of the scales used to assess cigarette dependence (i.e., NDSS and FTND) in NSDUH, an overall indicator of dependence was created. These two indicators were then combined to create an overall measure of cigarette dependence in the past month.

To create the NDSS indicator, responses to items CIGIRTBL, CIGCRAVE, CIGCRAGP, CIGAVOID, CIGPLANE, CIGRNOUT, CIGREGDY, CIGREGWK, CIGREGNM, CIGNOINF, CIGINCRS, CIGSATIS, and CIGLOTMR were coded from 1 to 5, where 1 = Not at all true, 2 = Somewhat true, 3 = Moderately true, 4 = Very true, and 5 = Extremely true. Responses to items CIGINCTL,

CIGNMCHG, CIGSVLHR, and CIGINFLU were reverse coded from 5 to 1, where 5 = Not at all true, 4 = Somewhat true, 3 = Moderately true, 2 = Very true, and 1 = Extremely true. For respondents who reported smoking cigarettes in the past month, the total NDSS score was calculated as the sum of the recoded responses to the previous 17 questions. A respondent was classified as dependent on cigarettes (nicotine) based on the NDSS (NDSSDEPND = 1) if the total NDSS score was greater than or equal to 46.75 (corresponding to an average score of 2.75), regardless of whether any of the 17 NDSS questions were missing. If a respondent provided a response to all 17 questions and the total NDSS score was less than 46.75, then the respondent was classified as not being dependent (NDSSDEPND = 0). If a respondent was missing at least 1 of the 17 questions, but the sum of the nonmissing items, plus five times the number of questions with missing data was less than 46.75, then the respondent was also classified as not dependent (NDSSDEPND = 0), because even the largest possible values for the missing responses would not sum to a score above the threshold. Otherwise, NDSSDEPND was set to missing for past month users of cigarettes. The average NDSS threshold of 2.75 was derived by the scale developers by examining the distribution of scores in other samples of smokers who were administered the NDSS, including a contrast of scores obtained for smokers who were not dependent (chippers) versus heavy smokers (Shiffman et al., 1994).

For the FTND measure, respondents who reported smoking cigarettes in the past month were classified as being dependent on cigarettes (nicotine) based on the FTND (FTNDDEPND = 1) if they first smoked a cigarette within 30 minutes of waking up on the days they smoked (CIGWAKE = 1 or 2). Past month users of cigarettes who reported smoking their first cigarette more than 30 minutes after waking up on the days they smoked (CIGWAKE = 3 or 4) were classified as not dependent (FTNDDEPND = 0). Otherwise, FTNDDEPND was set to missing for past month users of cigarettes.

Using both the NDSS and FTND measures, respondents were classified as being dependent on cigarettes (nicotine) (PMNICDEP = 1) if they met either the NDSS (NDSSDEPND = 1) or FTND (FTNDDEPND = 1) criteria for dependence. A past month cigarette user was classified as not dependent in the past month if NDSSDEPND = 0 and FTNDDEPND = 0. Otherwise, nicotine dependence status was unknown for past month cigarette users.

7.3.1.5 Creation of DSM-5 Substance Use Disorder Indicators

This section describes how the edited DSM-5 SUD indicators were created. The diagnosis of SUD is based on the presence of two or more criteria. The edited disorder indicators are created in the following three steps:

1. Membership in the domain is determined. Domain nonmembers (i.e., respondents for which past year SUD is not applicable) receive a skip code for the edited disorder indicator to indicate that they did not (mis)use the substance within the past year. Generally, the domain for nonprescription drugs includes past year users. For prescription drugs, two disorder indicators are created. For the first version, the domain includes past year misusers of a given substance. For the second version, the domain

includes past year users but not misusers of the substance. This step is covered in Section 7.3.1.5.1.

2. The item-level variables, described in Section 7.3.1.3, are recoded to criterion-level variables, where the criteria are those defined by DSM-5. This step is covered in Section 7.3.1.5.2.
3. The criterion-level variables are recoded to create disorder indicators defined by DSM-5. This step is covered in Section 7.3.1.5.3.

[Table 7.7](#) provides a quick reference guide to the variables used to create the DSM-5 disorder indicators. The 13 item-level variables used for all disorder indicators are UDxxLRGAMTS, UDxxNOTSTOP, UDxxWSHSTOP, UDxxTIMEGET, UDxxTIMEUSE, UDxxWANTBAD, UDxxSTRURGE, UDxxWORKPRB, UDxxFMLYCTD, UDxxSTOPACT, UDxxGETHURT, UDxxMNTLCTD, and UDxxHLTHCTD, where xx is the two-letter abbreviation associated with the substance. The item-level variables associated with tolerance that are used for all disorder indicators except those measuring disorder among past year users but not misusers of prescription drugs are UDxxNEEDMOR and UDxxLESSEFF. The item-level variable that is used for disorder indicators including withdrawal symptoms is derived from one or more withdrawal symptoms questions. Details appear in Section 7.3.1.5.2.

Table 7.7 Quick Reference Guide to Creation of Edited DSM-5 Disorder Indicators

Substance	Domain Variables	Item-Level Variables	Edited Disorder Indicator
Alcohol	ALCYR ALCYRTOT	Includes tolerance and withdrawal symptoms	PYUD5ALC
Marijuana	MRJYR MJYRTOT	Includes tolerance and withdrawal symptoms	PYUD5MRJ
Cocaine	COCYR CONDLREC	Includes tolerance and withdrawal symptoms	PYUD5COC
Heroin	HERYR HRNDLREC HRSNFREC HRSMKREC	Includes tolerance and withdrawal symptoms	PYUD5HER
Hallucinogens	HALLUCYR	Includes tolerance symptoms. Does not include withdrawal symptoms	PYUD5HAL
Inhalants	INHALYR	Includes tolerance symptoms. Does not include withdrawal symptoms	PYUD5INH
Methamphetamine	METHAMYR METHNDLRC	Includes tolerance and withdrawal symptoms	PYUD5MTH
Prescription Pain Relievers, Past Year Misusers	PNRNMYR	Includes tolerance and withdrawal symptoms	EDUD5PNRMIS
Prescription Tranquilizers, Past Year Misusers	TRQNMYS	Includes tolerance and withdrawal symptoms	EDUD5TRQMIS

Table 7.7 Quick Reference Guide to Creation of Edited DSM-5 Disorder Indicators (continued)

Substance	Domain Variables	Item-Level Variables	Edited Disorder Indicator
Prescription Stimulants, Past Year Misusers	STMNMYR	Includes tolerance and withdrawal symptoms	EDUD5STMMIS
Prescription Sedatives, Past Year Misusers	SEDNMYR	Includes tolerance and withdrawal symptoms	EDUD5SEDMIS
Prescription Pain Relievers, Past Year Users but Not Misusers	PNRANYR PNRNMYR	Does not include tolerance or withdrawal symptoms	EDUD5PNRUNM
Prescription Tranquilizers, Past Year Users but Not Misusers	TRQANYR TRQNMYR	Does not include tolerance or withdrawal symptoms	EDUD5TRQUNM
Prescription Stimulants, Past Year Users but Not Misusers	STMANYR STMNMYR	Does not include tolerance or withdrawal symptoms	EDUD5STMUNM
Prescription Sedatives, Past Year Users but Not Misusers	SEDANYR SEDNMYR	Does not include tolerance or withdrawal symptoms	UDUD5SEDUNM

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

7.3.1.5.1 Domain Variables for DSM-5 Disorder Indicators

For the SUD indicators, the membership in the domain is generally determined by the imputation-revised recency of use for alcohol and nonprescription drugs (i.e., marijuana, cocaine, heroin, hallucinogens, inhalants, and methamphetamine). For prescription drugs, two SUD indicators are created for each substance. The first indicator (e.g., EDUD5PNRMIS) is applicable to respondents who misused the drug in the past year and domain status is determined by the imputation-revised recency-of-misuse variable (e.g., PNRNMYR = 1). The second SUD indicator (e.g., EDUD5PNRUNM) for prescription drugs is applicable to respondents who reported use but not misuse in the past year and domain status is determined by the imputation-revised recency of any use and recency-of-misuse variables (e.g., PNRANYR = 1 and PNRNMYR = 0 for prescription pain relievers). In addition, the respondents who used alcohol or marijuana on 5 or fewer days in the past year were not administered the DSM-5 questions. As a result, respondents who reported using alcohol or marijuana on 5 or fewer days in the past year were excluded from the domain. For cocaine and methamphetamine, respondents could answer the disorder questions if they reported injecting the drug with a needle in the past year in the special drugs section, even if they reported not using the drug in the past year in the main questionnaire section associated with these drugs. Thus, respondents who reported injecting cocaine or methamphetamine with a needle in the past year were included in the domain. Similarly, respondents were administered the heroin use disorder questions if they reported injecting heroin with a needle or smoking, sniffing, or snorting heroin in the past year in the special drugs section, even if they had previously reported not using heroin in the past year in the heroin section of the questionnaire. Like for cocaine and methamphetamine, these respondents were included in the domain for past year heroin SUD.

Those who are not in the domain receive a skip code for the edited SUD indicators. Those who never used the drug receive a code of 91, and those who have used the drug in their lifetime but not in the past year receive a code of 93.

In summary, for hallucinogens and inhalants, the variable used to determine the domain status is the imputation-revised recency-of-use variable. The creation of these variables is described in Chapter 6.

For alcohol and marijuana, the respondent is in the domain if, and only if, both of the following is true:

- ALCYR/MRJYR is 1 (i.e., the respondent is a past year alcohol/marijuana user according to the imputation-revised recency variable).
- ALCYRTOT/MJYRTOT is greater than 5 or missing (i.e., the respondent reported using alcohol/marijuana on at least 5 days in the past 12 months or failed to provide a response to the 12-month frequency-of-use question).

For cocaine and methamphetamine, the respondent is in the domain if, and only if, at least one of the following is true:

- COCYR/METHAMYR is 1 (i.e., the respondent is a past year cocaine/methamphetamine user according to the imputation-revised recency variable).
- CONDLREC/METHNDLRC is 1 or 2 (i.e., the respondent used a needle to inject cocaine/methamphetamine in the past year).

For heroin, the respondent is in the domain if, and only if, at least one of the following is true:

- HERYR is 1 (i.e., the respondent is a past year heroin user according to the imputation-revised recency variable).
- HRNDLREC is 1 or 2, or HRSMKREC is 1 or 2, or HRSNFREC is 1 or 2 (i.e., the respondent used a needle to inject heroin or smoked, sniffed, or snorted heroin in the past year).

The variables CONDLREC, METHNDLRC, HRNDLREC, HRSMKREC, and HRSNFREC are described in Section 7.2.5.1 and do not undergo any formal imputation. This implies that the domain variable is technically missing if the past year use variable is not 1 and the question(s) from the special drugs section for a given substance are missing. For the sake of simplicity, and because this situation is rare, these respondents are assumed to be out of the domain.

For prescription drugs, the respondent is in the domain for the “misuse” SUD indicator if the following is true:

- PNRNMYR/TRQNMYR/STMNMYR/SEDNMYR is 1 (i.e., the respondent is a past year misuser of the prescription drug according to the imputation-revised recency variable).

For prescription drugs, the respondent is in the domain for the “use but not misuse” SUD indicator if, and only if, both of the following are true:

- PNRANYR/TRQANYR/STMANYR/SEDANYR is 1 (i.e., the respondent is a past year user of the prescription drug according to the imputation-revised recency variable).
- PNRNMYR/TRQNMYR/STMNMYR/SEDNMYR is 0 (i.e., the respondent is not a past year misuser of the prescription drug according to the imputation-revised recency variable).

7.3.1.5.2 *Creating the DSM-5 Criterion-Level Variables from the Item-Level Variables*

For hallucinogens and inhalants, the item-level variables are recoded into 10 criteria. For alcohol, marijuana, cocaine, heroin, and methamphetamine and the misuse of pain relievers, tranquilizers, stimulants, and sedatives, the item-level variables are recoded into 11 criteria. For use but not misuse of pain relievers, tranquilizers, stimulants, and sedatives, the item-level variables are recoded into 9 criteria. For substances that include withdrawal symptoms, respondents are asked whether they experienced substance-specific withdrawal symptoms at one time that lasted for longer than a day after they cut back or stopped using. The specific number and type of listed withdrawal symptoms varied by substance. The detailed descriptions of criteria for alcohol and illicit drugs are documented in the nicotine dependence and substance use disorder appendix of the 2023 Public Use File Codebook (Center for Behavioral Health Statistics and Quality, 2024b). [Table 7.8](#) provides a summary of item-level variables that are used to assess whether someone has an SUD based on DSM-5 diagnoses.

Table 7.8 DSM-5 Substance Use Disorder Diagnosis Criteria

DSM-5 Criteria	Criteria Description	Item-Level Variables
1	Used substance many times in larger amounts or for a longer time than intended	UDxxLRGAMTS=1
2	Unable to cut down or stop substance use or wished could cut down or stop	UDxxNOTSTOP=1 or UDxxWSHSTOP=1
3	Spent a great deal of time using substance, feeling its effects, or getting over the effects	UDxxTIMEGET=1 or UDxxTIMEUSE=1
4	Had a strong urge to use substance	UDxxWANTBAD=1 or UDxxSTRURGE=1
5	Had serious problems at work, school, or home	UDxxWRKPRB=1
6	Had problems with family or friends that were caused or made worse and continued to use substance even though it often caused arguments or problems with family or friends	UDxxFMLYPRD=1 and UDxxFMLYCTD=1
7	Gave up or spent a lot less time doing important social, occupational, or recreational activities	UDxxSTOPACT=1
8	Repeatedly got into situations where using substance increased chances of getting physically hurt	UDxxGETHURT=1

Table 7.8 DSM-5 Substance Use Disorder Diagnosis Criteria (continued)

DSM-5 Criteria	Criteria Description	Item-Level Variables
9	Continued to use substance even though it was causing problems with emotions, mental health, or physical health (Alcohol: or repeated blackouts)	UDxxMTLCTD=1 or UDxxHLTHCTD=1 (or UDALBLKCTD=1 for Alcohol)
10	Needed to use more substance to get the desired feeling, or using the same amount had less of an effect (Not applicable for past year users but not misusers of prescription drugs)	UDxxNEEDMOR=1 or UDxxLESSEFF=1
11	Reported experiencing substance withdrawal symptoms in the past year after substance use was cut back or stopped or used the same or similar substance to avoid having withdrawal symptoms (Not applicable for inhalants, hallucinogens, or past year users but not misusers of prescription drugs)	Alcohol, prescription tranquilizers, prescription sedatives: two or more (UDxxWDSWEAT=1, UDxxWDTRMBL=1, UDxxWDSLEEP=1, UDxxWDVOMIT=1, UDxxWDHALUC=1, UDxxWDSITST=1, UDxxWDFLANX=1, UDxxWDSZURE=1); or one or more (UDxxAVWALCO=1, UDxxAVWSVTR=1) Marijuana: three or more (UDMJWDANGRY=1, UDMJWDFLANX=1, UDMJWDSLEEP=1, UDMJWDAPPET=1, UDMJWDSITST=1, UDMJWDDEPRS=1) including at least one or more physical symptoms (UDMJWDSTMCH=1, UDMJWDSHAKE=1, UDMJWDSWEAT=1, UDMJWDFEVER=1, UDMJWDCHILL=1); or UDMJAVWMARJ=1 Cocaine, methamphetamine, prescription stimulants: UDxxWDFLBU=1 and two or more (UDxxWDTIRED=1, UDxxWDDREAM=1, UDxxWDSLEEP=1, UDxxWDHNGRY=1, UDxxWDSITST=1); or one or more (UDxxAVWCOCN=1, UDxxAVWMETH=1, UDxxAVWSTIM=1) Heroin, prescription pain relievers: three or more (UDxxWDFLBU=1, UDxxWDVOMIT=1, UDxxWDCRAMP=1, UDxxWDRUNNY=1, UDxxWDSWEAT=1, UDxxWDDIARR=1, UDxxWDYAWNS=1, UDxxWDFEVER=1, UDxxWDSLEEP=1); or one or more (UDxxAVWHERE=1, UDxxAVWP NRL=1)

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

7.3.1.5.3 Creating the Edited DSM-5 Disorder Indicators from the Criterion-Level Variables

Based on the DSM-5 definition of SUD, respondents are defined with disorder if, and only if, they are in the domain and have at least two positive criteria. Therefore, if p is the number of positive criteria and m is the number of missing criteria, the following can occur:

- The domain member is defined with disorder if, and only if, $p \geq 2$.
- The domain member is defined as not having a disorder if, and only if, $p + m < 2$.
- All other domain members have a missing value for the disorder indicator.

It follows that domain members have a missing value for the edited disorder indicator ([Table 7.7](#)) in the following scenarios:

- The domain member has no positive criteria and is missing two or more criteria.
- The domain member has one positive criterion and is missing one or more criteria.

7.3.1.6 Creation of Edited DSM-5 Substance Use Disorder Severity Variables

This section describes how the edited disorder severity variables are created under the DSM-5 diagnostic criteria. Respondents are defined with disorder when they have two or more positive criteria. SUDs are subclassified in terms of severity (i.e., mild, moderate, or severe). The severity of an SUD is determined by the number of individual criteria recorded as positive for a particular substance (or category of substances). An SUD is considered to be of mild severity when two or three criteria are present, of moderate severity when four or five symptoms are present, and severe when six or more symptoms are present ([Table 7.9](#)).

Table 7.9 DSM-5 Substance Use Disorder Diagnosis Determined by Number of Positive Criteria

SUD	SUD Severity	Number of Positive Criteria
No	None	0 to 1
Yes	Mild	2 to 3
Yes	Moderate	4 to 5
Yes	Severe	6 or more

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); SUD = substance use disorder.

The DSM-5 disorder severity variables are constructed from the edited criterion-level variables described in [Table 7.7](#). The edited severity variables were created in the following two steps:

1. Membership in the domain is determined. The domain for the edited severity variables includes respondents with an imputation-revised past year use disorder for the respective substance (e.g., IRPYUD5ALC = 1 for alcohol).
2. The criterion-level variables are recoded to create severity variables. If p is the number of positive disorder criteria and m is the number of missing disorder criteria, then the following can occur:
 - a. The domain member is defined as having a mild disorder if, and only if, $(p + m) \leq 3$.
 - b. The domain member is defined as having a moderate disorder if, and only if, $p > 3$ and $(p + m) \leq 5$.
 - c. The domain member is defined as having a severe disorder if, and only if, $p > 5$.
 - d. All other domain members have a missing value for the disorder severity variable.

For nonprescription substances, the edited DSM-5 disorder severity variables have the name PYSEV5xxx, where xxx is the three-letter abbreviation associated with the substance (ALC, MRJ, COC, HER, HAL, INH, or MTH). For prescription drugs, the edited DSM-5 disorder severity variables are named EDSEVxxxMIS and EDSEVxxxUNM, where xxx is the three-letter

abbreviation associated with the substance (PNR, TRQ, STM, or SED) and the suffix identifies the domain (MIS = past year misusers; UNM = past year users but not misusers).

7.3.2 Imputation of Cigarette Dependence and Substance Use Disorder Variables

The imputation procedures performed on the SUD variables are addressed in two sections. Section 7.3.2.1 describes the imputation process for past month cigarette dependence. Section 7.3.2.2 describes how the imputation-revised DSM-5 SUD indicator and severity variables were created from their edited predecessors.

7.3.2.1 Imputation of Cigarette Dependence

This section describes how the imputation-revised cigarette dependence indicator was created from the edited version of the same variable. The creation of the edited variable is described in Section 7.3.1.4. This variable underwent imputation using the modPMN method, which is described in Chapter 3.

The imputation procedures for cigarette dependence involve only a single variable. Because the imputation set contains only a single variable, no cycling is performed. Sections 7.3.2.1.1, 7.3.2.1.2, and 7.3.2.1.3 provide details on the three steps of modPMN as described in Chapter 3. Section 7.3.2.1.4 describes the variables created by these procedures—the imputation-revised dependence indicator and the imputation indicator.

7.3.2.1.1 Step 1: Response Propensity Adjustment

This section provides specific information about the response propensity step for cigarette dependence imputation. General information about the response propensity adjustment step of modPMN is available in Section 3.6.1. The four key inputs to the response propensity adjustment step, as laid out in Section 3.3.1, are described below.

- **Analysis Weight:** For cigarette dependence, the input to the response propensity model was the preliminary analysis weight.
- **Domain:** The domain includes past month cigarette users based on the imputation-revised recency-of-use variable (IRCIGRC).
- **Item Response Indicator:** This was based simply on whether the edited variable (PMNICDEP) was missing. The process by which this is determined is described in Section 7.3.1.4.
- **Covariates:** The covariates used in the response propensity model are summarized in Appendix D.

7.3.2.1.2 Step 2: Prediction Modeling

This section provides specific information about the prediction modeling step for cigarette dependence imputation. General information about the prediction modeling step of modPMN is provided in Section 3.6.2. The response propensity model segregated the domain members into five bins, and separate prediction models were fit within the bins.

For each of the prediction models, the probability of dependence was modeled using dichotomous logistic regression. SUDAAN's LOGISTIC procedure (called RLOGIST in SAS-callable SUDAAN) was used to estimate the regression coefficients and to calculate predicted probabilities for all item nonrespondents. These predicted probabilities were used in the subsequent step to impute a positive or negative response. The covariates used in the models are summarized in Appendix D.

7.3.2.1.3 Step 3: Collocated Stochastic Imputation

The collocated stochastic imputation step for the cigarette dependence indicator is straightforward and is described in general in Section 3.6.3. Briefly, each of the prediction models is followed by an imputation step where a random number is used to impute a value of "Yes" or "No" for the dependence indicator for each item nonrespondent. Imputation is done only at the dependence indicator level. This approach simplifies the procedures, but one shortcoming is that partial information is ignored. For example, all else being equal, an item nonrespondent who is known not to be dependent based on the FTND criteria but who is missing the NDSS criteria has the same probability of being imputed with dependence as an item nonrespondent who is missing the NDSS and FTND criteria. The numbers of respondents undergoing collocated stochastic imputation for cigarette dependence is provided in Appendix E.

7.3.2.1.4 Final Cigarette Dependence Variables

Once the imputation has been completed, the imputation-revised dependence variable (IRPMNICDEP) and the imputation indicator (IIPMNICDEP) are delivered to the project master dataset. The imputation indicator was assigned to three levels, where level 1 mapped to "Questionnaire data," level 2 mapped to "Logically assigned," and level 3 mapped to "Imputation-revised."

7.3.2.2 Imputation of DSM-5 Substance Use Disorder Indicators and Substance Use Disorder Severity Variables

This section describes how the imputation-revised DSM-5 SUD indicators and disorder severity variables were created from the edited versions of the same variables. The creation of the edited variables is described in Sections 7.3.1.5 and 7.3.1.6. These variables underwent imputation using the modPMN method, which is described in Chapter 3.

The DSM-5 SUD measures are created for 11 substances covering alcohol and illicit drugs. However, for the four prescription drugs, two sets of SUD measures are created so that a total of 15 imputation-revised disorder indicators and 15 imputation-revised disorder severity variables are created. Section 7.3.2.2.1 describes how these 30 variables were segregated into imputation sets, and Section 7.3.2.2.2 describes how the variables were ordered within the imputation sets. The processing order is important because the variables later in the sequence can sometimes seem to have undergone a more precise imputation than variables earlier in the sequence (see Section 3.7.1). Sections 7.3.2.2.3, 7.3.2.2.4, and 7.3.2.2.5 provide details on the three steps of modPMN as described in Chapter 3. Section 7.3.2.2.6 discusses how the three steps were arranged to complete imputations. Section 7.3.2.2.7 describes the 60 variables created by these procedures—the 30 imputation-revised disorder indicator and disorder severity

variables and their imputation indicators—and mentions some of the recoded variables created using these variables.

7.3.2.2.1 *DSM-5 Substance Use Disorder Indicators and Substance Use Disorder Severity: Imputation Sets*

Cocaine, methamphetamine, and the misuse of prescription stimulants were grouped into a set because cocaine and methamphetamine are stimulants. Heroin and the misuse of prescription pain relievers were considered to be closely related because both are opioids. Similarly, the misuse of tranquilizers and the misuse of sedatives were considered to be closely related. Indicators for use but not misuse of all prescription drugs were also grouped into a single set. This led to the formation of eight imputation sets for disorder and eight imputation sets for disorder severity as shown in [Table 7.10](#). Although these sets are numbered in the table, sets 1 through 7 and sets 9 through 15 can be run in any order (or in parallel) because none of the imputation-revised variables are used as covariates for any other variables outside of their own imputation set. However, sets 1 through 7 must be run before sets 8 through 15, and sets 8 through 15 must be run before set 16. Set 8 can be run in parallel with sets 9 through 15.

Table 7.10 Segregation of Substances into Imputation Sets

Substance	Substance Class	Imputation Set: Disorder	Imputation Set: Disorder Severity
Alcohol	Alcohol	Set 1	Set 9
Marijuana	Marijuana	Set 2	Set 10
Cocaine	Stimulants	Set 3	Set 11
Methamphetamine			
Misuse of Prescription Stimulants			
Heroin	Opioids	Set 4	Set 12
Misuse of Prescription Pain Relievers			
Hallucinogens	Hallucinogens	Set 5	Set 13
Inhalants	Inhalants	Set 6	Set 14
Misuse of Prescription Tranquilizers	Sedatives, Hypnotics, and Anxiolytics	Set 7	Set 15
Misuse of Prescription Sedatives			
Use but Not Misuse of Prescription Stimulants	Use but Not Misuse of Prescription Drugs	Set 8	Set 16
Use but Not Misuse of Prescription Pain Relievers			
Use but Not Misuse of Prescription Tranquilizers			
Use but Not Misuse of Prescription Sedatives			

7.3.2.2.2 Ordering of DSM-5 Variables within Imputation Sets

As explained in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rate. The order of imputation for sets with multiple variables in 2023 is shown in [Table 7.11](#).

Table 7.11 Sequencing of DSM-5 Disorder Indicator and Disorder Severity within Imputation Sets, 2023

Substance Class	Disorder Indicator Sequence	Disorder Severity Sequence
Stimulants	Cocaine, Stimulants, ¹ Methamphetamine	Cocaine, Methamphetamine, Stimulants ¹
Opioids	Heroin, Pain Relievers ¹	Heroin, Pain Relievers ¹
Sedatives, Hypnotics, and Anxiolytics	Sedatives, ¹ Tranquilizers ¹	Sedatives, ¹ Tranquilizers ¹
Use but Not Misuse of Prescription Drugs	Pain Relievers, Tranquilizers, Stimulants, Sedatives	Stimulants, Tranquilizers, Pain Relievers, Sedatives

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ Misuse of prescription drugs.

7.3.2.2.3 Step 1: Response Propensity Adjustment

This section provides specific information about the response propensity steps for the 16 disorder indicator and disorder severity imputation sets. General information about the response propensity adjustment step of modPMN is available in Section 3.6.1. The four key inputs to the response propensity adjustment step, as laid out in Section 3.3.1, are described below.

- **Analysis Weight:** For all imputation sets, the input to the response propensity model was the preliminary analysis weight.
- **Domain:** This is described in Sections 7.3.1.5.1 and 7.3.1.6. For the disorder indicator imputation sets, the domain includes past year (mis)users, or past year users but not misusers, based mainly on the imputation-revised recency of (mis)use variables. For disorder severity, the domain includes respondents with an imputation-revised past year use disorder for the respective substance.
- **Item Response Indicator:** This was based simply on whether the edited variable was missing. The process by which this is determined is described in Section 7.3.1.5.3 for the disorder indicators and in Section 7.3.1.6 for the disorder severity variables.
- **Covariates:** Information on the covariates used in each of the response propensity models are included in Appendix D.

7.3.2.2.4 Step 2: Prediction Modeling

This section provides specific information about the prediction modeling step of modPMN for DSM-5 SUD and disorder severity. General information about the prediction modeling step of modPMN is provided in Section 3.6.2. Each of the response propensity models segregated the domain members into five bins, and separate prediction models were fit within the bins.

For each of the disorder prediction models, the probability of disorder was modeled using dichotomous logistic regression. SUDAAN's LOGISTIC procedure was used to estimate the regression coefficients and to calculate predicted probabilities for all item nonrespondents. For each of the disorder severity prediction models, the probability of disorder severity was modeled using polytomous logistic regression. SUDAAN's MULTLOG procedure was used to estimate the regression coefficients and to calculate predicted probabilities for all item nonrespondents. These predicted probabilities were used in the subsequent step to determine the final imputed value. The covariates used in the prediction models are summarized in Appendix D.

7.3.2.2.5 Step 3: Collocated Stochastic Imputation

The collocated stochastic imputation step is straightforward and is described in general in Section 3.6.3. Briefly, each of the prediction models is followed by an imputation step where a random number is used to impute a value of "Yes" or "No" for the given disorder indicator or to impute a severity category for the given substance for each item nonrespondent.

Imputation is done only at the disorder or disorder severity level (i.e., individual criteria are not imputed). This approach simplifies the procedures, but one disadvantage is that partial information stored in the edited item-level variables (Sections 7.3.1.1, 7.3.1.2, and 7.3.1.3) and the edited criterion-level variables (Section 7.3.1.5.2) is not incorporated. For example, all else being equal, an item nonrespondent with one positive criterion and one missing criterion has the same probability of being imputed with disorder as an item nonrespondent with no positive criteria and two missing criteria. The numbers of nonrespondents undergoing collocated stochastic imputation for each of the 15 disorder indicators and 15 disorder severity variables are provided in Appendix E.

7.3.2.2.6 Sequencing of Steps for Selected Substance Use Disorder and Disorder Severity Imputation Sets

This section provides the sequencing of steps within an imputation set. General information about the sequencing of steps for modPMN is available in Sections 3.7.2 and 3.7.3.

Imputation sets 3 and 11 include three variables so there are six 12-step blocks of programs included (Section 3.7.2). The first of the three is for the first variable within the first cycle; the second is for the second variable within the first cycle; the third is for the third variable within the first cycle; and the fourth, the fifth, and the last are for the same ordered variables within the second cycle. Each 12-step block consists of a response propensity step where the domain members are segregated into five bins, first followed by a prediction modeling step within each bin, then followed by a collocated stochastic imputation step within each bin, and finally followed by a "wrap-up" program that merges the five datasets into one. Imputation sets 1, 2, 5, 6, 9, 10, 13, and 14 include only one variable so no cycling is performed. Imputation sets 4, 7, 12, and 15 include two variables each. Imputation sets 8 and 16 include four variables each. The algorithm for these sets is identical to the one used for imputation sets 3 and 11.

7.3.2.2.7 *Final DSM-5 Substance Use Disorder Indicators and Substance Use Disorder Severity Variables*

Once all the imputation sets have been completed, 60 variables are assembled for delivery to the project master dataset: the imputation-revised disorder indicator (IRPYUD5xxx for ALC, MRJ, COC, HER, HAL, INH, and MTH; IRUD5xxxMIS and IRUD5xxxUNM for PNR, TRQ, STM, and SED) and disorder severity variables (IRPYSEV5xxx for ALC, MRJ, COC, HER, HAL, INH, and MTH; IRSEVxxxMIS and IRSEVxxxUNM for PNR, TRQ, STM, and SED) for the 11 substances with SUD measures along with their imputation indicators.¹⁴³ The imputation indicators (IIPYUD5xxx/IIUD5xxxMIS/IIUD5xxxUNM and IIPYSEV5xxx/IISEVxxxMIS/IISEVxxUNM) were assigned to three levels, where level 1 mapped to "Questionnaire data," level 3 mapped to "Imputation-revised," and level 9 mapped to "legitimate skip." Respondents who were not in the domain for a particular outcome received values of 9 (legitimate skip).

Indicators of overall SUD (i.e., disorder, no disorder) were created for composite measures by combining the imputation-revised variables described previously. For example, a respondent is classified as having an opioid use disorder (UD5OPIMIS) if the respondent is classified as having a heroin use disorder (IRPYUD5HER=1) or the respondent is a past year misuser of pain relievers with a pain reliever use disorder (IRUD5PNRMIS=1). For specific prescription drugs (e.g., pain relievers), indicators were also created to combine the separate imputation-revised variables created for past year users and past year users but not misusers. [Table 7.12](#) lists the composite variables created using combinations of these imputed variables. These recoded variables are described in the nicotine dependence and substance use disorder appendix of the 2023 Public Use File Codebook (Center for Behavioral Health Statistics and Quality, 2024b).

Table 7.12 DSM-5 Recoded Variables

Variable Name	Variable Description
UD5TRQSDMIS	TRANQUILIZER OR SEDATIVE USE DISORDER - PAST YEAR MISUSERS
UD5TRQSDUNM	TRANQUILIZER OR SEDATIVE USE DISORDER - PAST YEAR UNM
UD5TRQSDANY	TRANQUILIZER OR SEDATIVE USE DISORDER - PAST YEAR USERS
UD5PSYMIS	PSYCHOTHERAPEUTIC USE DISORDER - PAST YEAR MISUSERS
UD5PSYUNM	PSYCHOTHERAPEUTIC USE DISORDER - PAST YEAR UNM
UD5PSYANY	PSYCHOTHERAPEUTIC USE DISORDER - PAST YEAR USERS
UD5OPIMIS	OPIOID USE DISORDER - PAST YEAR MISUSERS
UD5OPIUNM	OPIOID USE DISORDER - PAST YEAR UNM
UD5OPIANY	OPIOID USE DISORDER - PAST YEAR USERS
UD5HRPNRMIS	HEROIN AND/OR PAIN RELIEVER USE DISORDER - PAST YEAR MISUSERS
UD5HRPNRUNM	HEROIN AND/OR PAIN RELIEVER USE DISORDER - PAST YEAR UNM

¹⁴³ See Section 3.1 for details about imputation indicators.

Table 7.12 DSM-5 Recoded Variables (continued)

Variable Name	Variable Description
UD5HRPNRANY	HEROIN AND/OR PAIN RELIEVER USE DISORDER - PAST YEAR USERS
UD5ILLMIS	ILLCIT DRUG USE DISORDER - PAST YEAR MISUSERS
UD5ILLUNM	ILLCIT DRUG USE DISORDER - PAST YEAR UNM
UD5ILLANY	ILLCIT DRUG USE DISORDER - PAST YEAR USERS
UD5IEMMIS	ILLCIT DRUG OTHER THAN MARIJUANA USE DISORDER - PAST YEAR MISUSERS
UD5IEMUNM	ILLCIT DRUG OTHER THAN MARIJUANA USE DISORDER - PAST YEAR UNM
UD5IEMANY	ILLCIT DRUG OTHER THAN MARIJUANA USE DISORDER - PAST YEAR USERS
UD5ILALMIS	ILLCIT DRUG OR ALCOHOL USE DISORDER - PAST YEAR MISUSERS
UD5ILALUNM	ILLCIT DRUG OR ALCOHOL USE DISORDER - PAST YEAR UNM
UD5ILALANY	ILLCIT DRUG OR ALCOHOL USE DISORDER - PAST YEAR USERS
UD5ILAALMIS	ILLCIT DRUG AND ALCOHOL USE DISORDER - PAST YEAR MISUSERS
UD5ILAALUNM	ILLCIT DRUG AND ALCOHOL USE DISORDER - PAST YEAR UNM
UD5ILAALANY	ILLCIT DRUG AND ALCOHOL USE DISORDER - PAST YEAR USERS
UD5ILNALMIS	ILLCIT DRUG USE DISORDER, NO ALCOHOL USE DISORDER - PAST YEAR MISUSERS
UD5ILNALUNM	ILLCIT DRUG USE DISORDER, NO ALCOHOL USE DISORDER - PAST YEAR UNM
UD5ILNALANY	ILLCIT DRUG USE DISORDER, NO ALCOHOL USE DISORDER - PAST YEAR USERS
UD5ALNLMIS	ALCOHOL USE DISORDER, NO ILLCIT DRUG USE DISORDER - PAST YEAR MISUSERS
UD5ALNILUNM	ALCOHOL USE DISORDER, NO ILLCIT DRUG USE DISORDER - PAST YEAR UNM
UD5ALNILANY	ALCOHOL USE DISORDER, NO ILLCIT DRUG USE DISORDER - PAST YEAR USERS
UD5CNSMIS	CENTRAL NERVOUS SYSTEM STIMULANT USE DISORDER - PAST YEAR MISUSERS
UD5CNSUNM	CENTRAL NERVOUS SYSTEM STIMULANT USE DISORDER - PAST YEAR UNM
UD5CNSANY	CENTRAL NERVOUS SYSTEM STIMULANT USE DISORDER - PAST YEAR USERS
UDYR5PNRANY	PAIN RELIEVER USE DISORDER - PAST YEAR USERS
UDYR5SEDANY	SEDATIVE USE DISORDER - PAST YEAR USERS
UDYR5STMANY	STIMULANT USE DISORDER - PAST YEAR USERS
UDYR5TRQANY	TRANQUILIZER USE DISORDER - PAST YEAR USERS
SEVYRPNRANY	PAIN RELIEVER USE DISORDER SEVERITY - PAST YEAR USERS
SEVYRSEDANY	SEDATIVE USE DISORDER SEVERITY - PAST YEAR USERS
SEVYRSTMANY	STIMULANT USE DISORDER SEVERITY - PAST YEAR USERS
SEVYRTRQANY	TRANQUILIZER USE DISORDER SEVERITY - PAST YEAR USERS

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); UNM = users but not misusers.

8. Editing and Imputation in Additional Sections for Substance Use

8.1 Introduction

This chapter discusses editing and imputation procedures for the following additional sections that pertain to substance use:

- blunts,
- prior substance use,
- alcohol and drug use treatment,
- consumption of alcohol,
- emerging issues, and
- market information for marijuana.

Most edited variables described in this chapter for alcohol and drug use treatment were imputed to remove missing values. Imputed variables in the emerging issues section were also imputed for kratom, vaping of only flavoring, illegally made fentanyl (IMF), synthetic marijuana, and synthetic stimulants. Section 8.4.6 describes the imputation procedures for substance use treatment. Imputation of variables from the emerging issues section is covered in Section 8.6.6.

Discussion of interview sections in this chapter is based on related topical content rather than the sequential order in which these sections appeared to respondents during the interview. The order in which these sections were presented to respondents relative to other sections of the interview can be found in the computer-assisted interviewing (CAI) specifications for the 2023 National Survey on Drug Use and Health (NSDUH) that are available on the [NSDUH Data Collection](#) web page.

8.2 Blunts

Respondents were asked if they ever smoked part or all of a cigar with marijuana in it, sometimes called a “blunt.” If respondents were lifetime users of blunts, they were asked how long it had been since they had smoked a blunt. If respondents indicated blunt use in the past 30 days, they were asked to report the number of days they smoked blunts in the past 30 days. If respondents reported use of both cigars (from the tobacco section) and blunts in the past 30 days, they also were asked if they had smoked a cigar without marijuana in it during the past 30 days.

In addition, the section included consistency checks in selected situations where respondents gave reports of blunt use that were inconsistent with what they had reported in the marijuana section. Specifically, respondents could have reported earlier in the interview that they never used marijuana or any cannabis product and then indicate that they had smoked a blunt in the past 30 days. Similarly, respondents could have reported that their last use of marijuana was more than 30 days ago and then indicate use of blunts in the past 30 days. In these situations where respondents reported smoking blunts in the past 30 days but did not report marijuana use

in that period, respondents were asked to indicate which answer was correct (i.e., use of blunts in the past 30 days or their previous answer for marijuana use). If respondents confirmed smoking all or part of a cigar with marijuana in it in the past 30 days, they were asked why they reported earlier that they had never used marijuana or any cannabis product or that they had not used it in the past 30 days.

An important aspect of the processing of variables in this section pertaining to use of blunts consisted of assigning codes of 91, 93, and 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents said they never smoked part or all of a blunt, they were skipped out of the remaining blunts questions. Thus, respondents were assigned a code of 91 (NEVER USED BLUNTS) to the variables BLNTREC (most recent use of a blunt, corresponding to question BL02) and BLNT30DY (frequency of use of blunts in the past 30 days, corresponding to question BL02A). Similarly, if respondents refused to report whether they had ever smoked part or all of a blunt, that refusal code was assigned to BLNTREC and BLNT30DY as well. In addition, if respondents reported being lifetime users of blunts but reported that they last used blunts more than 30 days ago, question BL02a would have been skipped. Therefore, BLNT30DY was assigned a code of 93 (DID NOT USE BLUNTS IN THE PAST 30 DAYS) in this situation.

Minimal editing was done to the variables BLNT30C1, BLNT30C2, RSNOMRJ, and RSNMRJMO, corresponding to questions BL03, BL04, BL05, and BL06, respectively. These variables pertained to the above-mentioned consistency checks between respondents' answers in the blunts and marijuana sections. If respondents reported not using blunts in the past month or reported that they were a lifetime marijuana user, questions BL03 and BL05 would have been skipped. In this situation, BLNT30C1 and RSNOMRJ were assigned a code of 99. Similarly, if respondents were not past month users of blunts or reported last using marijuana in the past 30 days in the marijuana section (i.e., a response in the marijuana section that was consistent with reported use of blunts in the past 30 days), then questions BL04 and BL06 would have been skipped. In these situations, BLNT30C2 and RSNMRJMO were assigned a code of 99.

If respondents were routed to question BL03 (because they reported past month use of blunts but previously reported never using marijuana) and then they refused to answer question BL03, a refusal was propagated from BLNT30C1 to RSNOMRJ. If respondents answered BL03 as "don't know," RSNOMRJ retained a code of 98. Similarly, if respondents were routed to BL04 (because they reported past month use of blunts but previously reported using marijuana but not in the past 30 days) and they refused to answer, a refusal was propagated from BLNT30C2 to RSNMRJMO. Respondents answering "don't know" to question BL04 retained a code of 98 in RSNMRJMO.

Respondents who were lifetime users of blunts also were asked how old they were when they first used blunts (BL08). If respondents reported first using blunts within 1 year of their current age, they were asked to report the specific year and month when they first used, with the response options of 2 years before the current year, 1 year before the current year, and the current year (i.e., 2021, 2022, and 2023 in the 2023 questionnaire). If respondents reported first using blunts at their current age, their last birthday was in the current year, and they were

interviewed after their last birthday, the CAI program assumed that the first use of blunts occurred in the current year. These respondents were asked only for the month when they first used in the current year. The remaining respondents who first used blunts within 1 year of their current age could be routed to one of two possible questions on the specific year they first used based on the relationship between their date of birth and when the interview took place. Once respondents reported the year when they first used, they were asked to report the specific month they first used blunts in the year they had reported previously.

Because the routing logic to the different versions of the month of first use (MFU) and year of first use (YFU) questions was mutually exclusive, composite sets of MFU and YFU variables (BLNTYFU and BLNTMFU, respectively) were created from the individual unedited variables. In addition, the final YFU variables were recoded to replace codes of 1 or 2 from the questionnaire with values for the years (e.g., 2021 through 2023). In situations in which the CAI program skipped the YFU questions because it could be logically inferred that first use of blunts occurred in the interview year, a code matching the current year was assigned to BLNTYFU. This edit was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the MFU and YFU questions had been skipped because respondents first used blunts at ages that were more than 1 year below their current ages, legitimate skip codes were assigned to the BLNTYFU and BLNTMFU variables.

A consistency check also was included for situations in which the apparent age when respondents first used blunts that was calculated from their MFU and YFU and their date of birth (CAI variable MYR1STBL) was inconsistent with the age that respondents directly reported for when they first used blunts (CAI variable AGE1STBL). For example, the consistency check was triggered if 16-year-old respondents reported first using blunts at age 16 but then reported first using them in a year and month that would have meant they were 15 years old when they first used blunts. No further editing needed to be done if respondents indicated twice in a row that the AFU that was calculated from the MFU and YFU was correct. The CAI program updated the value for AGE1STBL (corresponding to the edited variable BLNTAGE) to agree with the values for the MFU and YFU.

The CAI program did not calculate an AFU based on respondents' reported MFU and YFU of blunts if respondents reported that they first used blunts in the same month as their birth month. In this situation, a unique AFU could not be determined. Consequently, the consistency check was not triggered in this situation. Similarly, the consistency check was not triggered if respondents had missing data in either of the year or month questions, such as if respondents knew the year when they first used blunts but they did not know the MFU.

If respondents indicated at some point in the consistency check sequence that the value they had originally reported for their AFU was correct (and by extension, that the MFU and YFU was not correct), they had an opportunity to revise the values for their MFU and YFU. These revised reports for MFU and YFU were used in subsequent editing steps. Otherwise, respondents' answers to the original MFU and YFU questions were used in subsequent editing.

[Table B.13](#) in Appendix B discusses edits pertaining to the consistency checks for these initiation variables for blunts. For consistency with how the substance use initiation data have been edited, the default when a respondent did not resolve an inconsistency between the AFU and the MFU and YFU was to favor the AFU in subsequent editing decisions. Unlike the initiation data in the substance use sections at the beginning of the interview, however, the initiation variables for blunts did not undergo subsequent statistical imputation.

Because these initiation variables were added to the blunts section after the section was first included in the NSDUH questionnaire, editing of BLNTREC did not take into account these initiation data to keep the creation of BLNTREC comparable across years. However, a flag variable (BLRECFLG) has been included to indicate when initiation data suggested more recent use than what respondents indicated in BLNTREC. The flag was coded as 98 (blank) if no inconsistency was present between BLNTREC and the initiation data.

[Table B.14](#) presents additional edits pertaining to variables in the blunts section. For example, if respondents reported using blunts but they also reported never using marijuana, then question BL03 was triggered only when respondents reported using blunts in the past 30 days. Respondents who verified that they had never used marijuana would then have their BLNTEVER answer assigned to 4 (No LOGICALLY ASSIGNED) and BNLTRC assigned to 81 (NEVER USED BLUNTS, Logically assigned). [Table B.14](#) also discusses miscellaneous edits for the initiation variables other than those described in [Table B.13](#) that were related to consistency checks.

No editing was done to make the data for blunts consistent with initiation data for cigars or marijuana from their respective earlier sections of the interview, or vice versa. Thus, for example, initiation data in the blunts section could indicate that respondents first smoked a cigar with marijuana in it at an earlier age than they reported for when they first used cigars or marijuana. However, variables in the blunts section of NSDUH codebooks included a standard footnote to indicate that the data in this section may be inconsistent with data from earlier sections of the interview.

8.3 Prior Substance Use

The prior substance use section covered the following topics:

- retrospective use of marijuana, cigarettes, alcohol, and cocaine in the year prior to the past 12 months (also referred to below as the past 12 to 24 months); and
- the age, year, and month when respondents last used substances covered in the corresponding substance use section of the interview (i.e., age at last use [ALU], month of last use [MLU], and year of last use [YLU]) if they were lifetime but not past month users of these substances.

One of the important aspects of the processing of variables in this section consisted of assigning codes of 91 and 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents never used marijuana or indicated use in the past 30 days, they were skipped out of the questions asking for their age, year, and month when they last used marijuana. Respondents were skipped out of the questions pertaining to the age,

year, and month when they last smoked cigarettes daily if (1) they smoked every day in the past 30 days or (2) they never had a period in their lives when they smoked cigarettes every day for at least 30 days. When a recency variable from a given substance use section (e.g., MJREC for marijuana) had a refusal code, and the corresponding variables in the prior substance use section had been skipped, that refusal from the substance use recency was propagated to the edited prior substance use variables (Section 2.4.3). Documentation of edits for specific subsections of the prior substance use section is presented in the remainder of this section.

8.3.1 Retrospective Use

Questions about retrospective use pertained to the use of marijuana, cigarettes, alcohol, and cocaine in the year before the past 12 months (i.e., in the period between 12 and 24 months before the interview). The corresponding edited variables were MRJYRBFR, CIGYRBFR, ALCYRBFR, and COCYRBFR, respectively.

If respondents never used the drug of interest (e.g., marijuana), the retrospective variable (e.g., MRJYRBFR) was assigned a code of 91. In addition, the cigarette recency variable CIGREC indicated if respondents last smoked cigarettes more than 3 years ago (CIGREC = 4). Logically, if respondents last smoked cigarettes more than 3 years ago, they would not have smoked a cigarette in the past 12 to 24 months. Therefore, when CIGREC indicated that respondents last smoked cigarettes more than 3 years ago and question LU37 (corresponding to CIGYRBFR) had been skipped, CIGYRBFR was assigned a code of 99 (LEGITIMATE SKIP).

These retrospective variables were not edited for consistency with other data for these drugs in the prior substance use section. For example, if MRJYRBFR indicated that respondents used marijuana in the past 12 to 24 months but respondents indicated that they last used marijuana at an age in edited variable MRJAGLST that would suggest that they last used marijuana more than 2 years ago, no editing was done to either MRJYRBFR or MRJAGLST. However, a codebook "NOTE" was added to alert analysts that these types of inconsistencies could exist between the related variables for these drugs.

8.3.2 Last Use of Drugs

The prior substance use section included questions about the last use of cigarettes, smokeless tobacco, cigars, alcohol, marijuana, cocaine (including crack cocaine if applicable), heroin, hallucinogens (including lysergic acid diethylamide [LSD], phencyclidine [PCP], and Ecstasy, if applicable), inhalants, or methamphetamine. If respondents were lifetime users of specific substances in the corresponding substance use section but had not used these substances in the past 30 days, they were asked in this section for the age, year, and month when they last used these substances (i.e., ALU, YLU, and MLU). If respondents ever had a period of smoking cigarettes daily but had not smoked every day in the past 30 days, they also were asked for the age, the year, and the month when they last smoked cigarettes on a daily basis.

Respondents who last used a given substance more than 30 days ago¹⁴⁴ were asked how old they were when they last used that substance. If respondents reported last using the substance within 1 year of their current age, they were asked to report the specific month and year when they last used, with response options for 2 years before the current year, 1 year before the current year, and the current year. If respondents reported last using the substance at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the last use of the substance occurred in the current year. These respondents were asked only for the month that they last used in the current year. The remaining respondents who last used a substance within 1 year of their current age could be routed to one of two possible questions on the specific year they last used. They were then routed to a question to report on the specific month that they last used the substance in the year they had reported previously.

Because the routing logic to the different versions of the MLU and YLU questions was mutually exclusive, composite sets of MLU and YLU variables were created from the individual unedited variables. In addition, if respondents indicated a specific year that they last used a substance, the final YLU variables were recoded to replace codes of 1 or 2 from the questionnaire with values for the years (e.g., 2021 through 2023). If respondents confirmed that they last used a substance at their current age and were interviewed subsequent to their birthday, a code matching the current year was assigned to the YLU. This edit was done even if respondents did not know what month they last used in the current year, or if they refused to report what month they last used in the current year. If the MLU and YLU questions had been skipped because respondents last used the substance more than 1 year younger than their current ages, legitimate skip codes were assigned to the final MLU and YLU variables.

Consistency checks included in this section were triggered when the values for the MLU and YLU were inconsistent with the ALU. Specifically, for respondents who recently stopped use of a given substance, the CAI program calculated a second ALU based on the MLU and YLU data by comparing these data with the respondent's date of birth. This comparison was not done if respondents reported last use of the substance in their birth month. A unique ALU could not be determined from the MLU and YLU in these situations because it was not known whether the substance use occurred before or after these respondents' birthdays. Similarly, a consistency check was not triggered if respondents had missing data in either of the month or year questions, such as if respondents knew the year when they last used a substance but did not know the MLU.

In the remaining situations in which respondents provided complete data for the MLU and YLU, a consistency check was triggered if the MLU and YLU suggested that respondents stopped use of the substance at an earlier or a later age than what they had previously reported in their ALU question. For example, a consistency check was triggered if 16-year-old respondents reported last using a substance at age 16 but then reported last using the substance in a month and year that would have meant the respondents were 15 years old when they last used it. No

¹⁴⁴ Subsequent discussion also applies to respondents whose last period of smoking cigarettes every day occurred more than 30 days ago.

editing needed to be done if respondents indicated twice in a row that the ALU that was calculated from the MLU and YLU was correct. The CAI program updated the value for the ALU (e.g., AGELSTCG for cigarettes) to agree with the values for the MLU and YLU.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their ALU (e.g., question LU03 for cigarettes) was correct, they had an opportunity to revise the values for their MLU and YLU. If a consistency check was triggered between the ALU and data in the MLU and YLU, the MLU and YLU were updated with any year and month data that the respondent entered in the consistency checks (e.g., LUCG07 and LUCG07a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the MLU and YLU data were picked up from the original source variables (e.g., LU03a through LU03d for any cigarette use) for use in subsequent editing.

[Table B.15](#) presents the edits that were implemented when consistency checks were triggered between the ALU, MLU, and YLU. The default when a respondent did not resolve an inconsistency between the ALU and the MLU and YLU was to favor the ALU in subsequent editing decisions.

Respondents also were skipped out of the MLU and YLU questions if they indicated that they last used a substance (or last smoked cigarettes every day) at an age that was more than 1 year younger than their current age. In these situations, the edited MLU and YLU variables were assigned legitimate skip codes.

Prior substance use variables were not edited with respect to *imputed* substance use variables. Suppose, for example, that respondents did not know or refused to report when they first used marijuana, but they gave ages at last use for marijuana in question LU02 that were consistent with their current ages. Although the imputed marijuana AFU (IRMJAGE) could have been imputed to a value greater than the ALU, the marijuana ALU (MRJAGLST) was not edited for consistency with IRMJAGE.

Similarly, if a substance's recency variable (e.g., MJREC for marijuana) had been assigned an "indefinite" value of 8 or 9 (indicating use at some point in the past 12 months or lifetime, respectively; see Section 6.2.2.3), the corresponding imputed recency (e.g., IRMJRC for marijuana) could be statistically imputed to indicate past month use (IRMJRC = 1). Although the prior substance use section was designed for respondents who were not past month users, data in this section were retained for respondents who were statistically imputed to be past month users. In this situation, analysts would have the option of deciding whether to use or disregard data from respondents who were imputed to be past month users.

[Table B.16](#) describes the specific edits that were implemented for the ALU, MLU, and YLU variables in the prior substance use section (i.e., other than the edits described in [Table B.15](#) when consistency checks were triggered between the ALU, MLU, and YLU variables). For example, a consistency check was triggered if respondents entered an ALU that was earlier than the age when they reported first using drugs or cigarettes, or when they first smoked cigarettes daily. Because these ALU questions occurred in a section of the interview after the questions that were used to identify past year initiates, respondents were not allowed to change their

answers to the corresponding AFU questions from the earlier substance use sections. Thus, the only way that respondents could resolve the inconsistency between the ALU and AFU was to change their answer to the ALU question. If respondents indicated that their inconsistent ALU was correct or they entered a new ALU value that was still inconsistent with the AFU, the edited ALU variable (e.g., MRJAGLST for marijuana) was assigned a bad data code. Thus, the relevant AFU from the corresponding earlier substance use section was used as the standard against which the corresponding ALU variable was compared. Similarly, the MFU and YFU questions for a given drug from the earlier substance use section of the interview were used as standards for editing the MLU and YLU variables for that drug.

8.4 Alcohol and Drug Use Treatment

In 2022, the section on alcohol and drug use treatment replaced the drug use treatment section from prior years. The new questionnaire section asked respondents about the following:

(1) receipt of specific sources of inpatient or outpatient treatment received in the past 12 months for alcohol or drug use (not including nicotine or tobacco); (2) receipt of treatment in the past 12 months for alcohol or drug use using prescription medication (also called medication-assisted treatment [MAT]), treatment via telehealth (i.e., over the phone or through video), or treatment in a prison, jail, or juvenile detention center; (3) substances for which treatment was received; (4) receipt of services in the past 12 months from other sources; (5) perceived unmet need for treatment; and (6) reasons for not receiving treatment in the past 12 months for alcohol or drug use.

Respondents were routed to the alcohol and drug use treatment section if they had reported previously in the survey that they had ever used alcohol, marijuana, cocaine or crack, heroin, hallucinogens, inhalants, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives. Therefore, all edited substance treatment variables were assigned a code of 91 or 9991 (NEVER USED ALCOHOL OR DRUGS) if respondents were skipped out of the entire alcohol and drug use treatment section because they never used alcohol or the drugs noted previously. If respondents were skipped out of the alcohol and drug use treatment questions because it was unknown whether they ever used alcohol or drugs, then all edited substance use treatment variables retained a code of 98 or 9998 (BLANK).

8.4.1 Type and Location of Alcohol and Drug Treatment

Sources of inpatient treatment in the past 12 months for alcohol or drug use that were asked about included (1) a hospital, (2) a residential drug or alcohol rehab or treatment center, (3) a residential mental health treatment center, and (4) some other place where respondents stayed overnight or longer. Sources of outpatient treatment in the past 12 months for use of alcohol or drugs that were asked about included (1) a drug or alcohol treatment or rehab center; (2) a mental health treatment center; (3) the office of a therapist, psychologist, psychiatrist, or substance use treatment professional; (4) a general medical clinic or doctor's office; (5) a hospital; (6) a school health or counseling center; and (7) some other place as an outpatient. Respondents also were asked whether they took medication in the past 12 months that was

prescribed to them to help cut back or stop alcohol use or drug use, whether they received treatment in this period for alcohol or drug use via telehealth (over the phone or through video), or whether they received treatment in a prison, jail or juvenile detention center.

Respondents who indicated that they received treatment at some other place where they stayed overnight or longer were asked in "OTHER, Specify" question TXSBIPDO to describe the other place where they received treatment. Similarly, respondents who indicated that they received treatment at some other place as an outpatient were asked in "OTHER, Specify" question TXSBOPGO to describe the other place where they received outpatient treatment. Principle for coding these "OTHER, Specify" variables are discussed in Section 2.3.1.3. If respondents reported that they did not receive inpatient or outpatient treatment in "some other place," the respective inpatient or outpatient edited "OTHER, Specify" variable (SUTINOPLSP or SUTOUTOPLSP) was assigned a legitimate skip code.

Subsequent sections for locations where respondents received alcohol or drug use treatment or other alcohol or drug use services discuss edits according to respondents' answers in the "OTHER, Specify" questions. [Table B.17](#) discusses these edits in more detail.

8.4.1.1 Editing of Inpatient and Outpatient Treatment Variables

Consistent with the procedures described in Section 2.4.5, if respondents did not report that they received inpatient treatment in the past 12 months in a particular location in questions TXSBIPTA to TXSBIPTC, but their response to question TXSBIPDO was coded to indicate that respondents received inpatient treatment in the past 12 months in one of these previously mentioned locations, then the corresponding edited variable for that location was assigned a code of 3 (Yes LOGICALLY ASSIGNED). For example, if respondents did not report that they received inpatient treatment in a hospital, but they specified "hospital" as the other place where they stayed overnight to receive alcohol or drug use treatment, then they were logically inferred to have received treatment in a hospital as an inpatient in the past 12 months. Similarly, if respondents did not report that they received outpatient treatment in the past 12 months in a particular location in questions TXSBOPTA to TXSBOPTF, but they specified that location in question TXSBOPGO as the other place where they received outpatient treatment in the past 12 months, then the corresponding edited variable for that location was assigned a code of 3.

In addition, respondents could have reported that they received alcohol or drug use treatment in some other place as an "outpatient" in the past 12 months but specified a residential location where they would be expected to stay overnight. For these respondents, the edited variable SUTINOPL for other inpatient locations was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTOUTOPLSP)] to indicate the editing across the inpatient and outpatient data. Similar edits for respondents who specified outpatient locations as other "inpatient" locations where they received alcohol or drug use treatment were in place.

8.4.1.2 Editing of Variables for Treatment Received as MAT, via Telehealth, or in a Prison, Jail, or Juvenile Detention Center

Respondents could specify in questions TXSBIPDO or TXSBOPGO that the other location where they received “inpatient” or “outpatient” treatment was via telehealth; in a prison, jail, or juvenile detention center; or at a dedicated MAT facility for opioid use (e.g., a methadone clinic). In these situations, the edited variables SUTPHVID (treatment by phone or video), SUTPRISON (treatment in a prison, jail, or juvenile detention center), or SUTRXDRG (use of prescription medication to cut back or stop drug use¹⁴⁵) were assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP or SUTOUTOPLSP)].

Respondents were routed to question TXALRX regarding the use in the past 12 months of prescription medication to cut back or stop alcohol use only if they had previously reported ever drinking alcohol. Therefore, respondents who never used alcohol but were eligible to be asked the substance use treatment questions because they reported lifetime use of a drug were skipped out of TXALRX. These respondents were assigned a code of 93 (USED DRUGS BUT NOT ALCOHOL) to the edited variable SUTRXALC.

Similarly, respondents were routed to question TXDRRX regarding the use in the past 12 months of prescription medication to help cut back or stop drug use only if they had previously reported ever using heroin or taking prescription pain relievers. Therefore, respondents who never used heroin or prescription pain relievers but were eligible to be asked the substance use treatment questions because they reported lifetime use of alcohol or a drug other than heroin or prescription pain relievers were skipped out of TXDRRX. These respondents were assigned a code of 93 (USED ALCOHOL OR DRUGS BUT NOT HEROIN OR PAIN RELIEVERS) to the edited variable SUTRXDRG.

8.4.2 Treatment for Specific Substances

Respondents who reported that they received substance use treatment at any inpatient location or any outpatient location in the past 12 months were asked follow-up questions regarding which specific substances for which they received treatment. Respondents who reported that they received substance use treatment via telehealth; in a prison, jail, or juvenile detention center; or as MAT for alcohol or drug (i.e., opioid) use in the past 12 months were asked follow-up questions regarding whether treatment was received for alcohol use only, for drug use only, or for both alcohol and drug use.

8.4.2.1 Editing of Variables for Inpatient or Outpatient Treatment for Specific Substances

A key aspect of editing the variables for inpatient or outpatient treatment for the use of specific substances involved assigning codes when respondents were not asked the questions. These

¹⁴⁵ Although question TXDRRX (corresponding to edited variable SUTRXDRG) asked about medication prescribed to respondents to help them cut back on or stop their “drug use,” respondents were asked this MAT question only if they reported lifetime use of opioids (i.e., heroin or prescription pain relievers). Hence, TXDRRX applies specifically to MAT for opioid use.

edits differed according to whether respondents reported or did not report that they received inpatient or outpatient treatment. Additional edits were implemented if respondents reported that they received inpatient or outpatient treatment.

8.4.2.1.1 *Editing for Respondents Who Received Inpatient or Outpatient Treatment*

Respondents who reported that they received substance use treatment at any inpatient or outpatient location were asked about the specific substances (e.g., marijuana) for which they received inpatient or outpatient treatment, respectively. Respondents were asked about treatment for use of the specific substances for which they had previously reported use in their lifetime. For example, respondents who reported that they received treatment at an inpatient location and reported lifetime use of alcohol and pain relievers were asked whether they received inpatient treatment for their use of alcohol (TXSBSUIA) and for their use of prescription pain relievers. If respondents in this example were not asked questions about their receipt of inpatient or outpatient treatment for the use of substances that they never used in their lifetime, then the edited variables corresponding to these substances were assigned a code of 93 (USED ALCOHOL OR DRUGS BUT NOT *[fill in name of specific drug]*). For example, if respondents received inpatient treatment but never used heroin, then the edited variable SUTINHER for the receipt of inpatient treatment for the use of heroin was assigned a code of 93 (USED ALCOHOL OR DRUGS BUT NOT HEROIN). However, if respondents were not asked a question about inpatient or outpatient treatment for their use of a substance because it was unknown whether they ever used that substance, then the corresponding variable for that substance retained a code of 98 (BLANK). For example, if respondents received inpatient treatment but they did not know or refused to report whether they ever used heroin, then SUTINHER retained a code of 98.

In addition, all respondents who reported inpatient or outpatient treatment were asked whether they received treatment for “some other drug.” Hence, codes of 93 were not assigned to the edited variables for the receipt of inpatient or outpatient treatment for the use of some other drug (edited variables SUTINODG and SUTOUTODG, respectively).

Respondents who indicated that they received treatment as an inpatient for the use of some other drug were asked in “OTHER, Specify” question TXSBIPSP to type in the name of the other drug for which they received treatment. Similarly, respondents who indicated that they received treatment for some other drug as an outpatient were asked in “OTHER, Specify” question TXSBIPSP to type in the name of the other drug for which they received treatment. Principles for coding these “OTHER, Specify” variables are discussed in Section 2.3.1.1.2. If respondents reported that they did not receive inpatient or outpatient treatment for the use of “some other drug,” then the respective inpatient or outpatient edited “OTHER, Specify” variables (SUTINODGSP1 through SUTINODGSP5 or SUTOUTODGSP1 through SUTOUTODGSP5) were assigned a legitimate skip code.

[Table B.17](#) presents edits that were specific to the variables for the receipt of inpatient or outpatient treatment for the use of specific substances. For example, if respondents did not report that they received inpatient treatment for the use of a particular substance during the

past 12 months, but they specified this substance as “some other drug” for which they received inpatient treatment, then the corresponding edited variable for that substance was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from SUTINODGSP1-SUTINODGSP5)]. If respondents specified in the “OTHER, Specify” question that they received inpatient treatment for a substance that they did not previously report that they ever used, then the corresponding edited variable for that substance was assigned a code of 13 [Yes LOG ASSGN (from SUTINODGSP1-5 but no lifetime use)].

In addition, some respondents did not report treatment for the use of any of the specific substances that they used in their lifetime and that they did not receive treatment for their use of some other drug. The following edits were implemented when this pattern occurred:

- Respondents who reported lifetime use of only one substance and who never used other substances were inferred to have received treatment for the use of the substance they used in their lifetime.
 - For these respondents, a code of 7 (Yes LOGICALLY ASSIGNED) was assigned to the corresponding edited substance-specific variable.
 - For example, respondents who reported that they received inpatient treatment, used only alcohol in their lifetime, and never used other substances were assigned a value of SUTINALC=7.
- Respondents who reported lifetime use of more than one drug but reported that they never used alcohol were assigned a code of 17 (Received treatment for unspecified drug) to the edited variable corresponding to treatment for other drug use (SUTINODG or SUTOUTODG for inpatient or outpatient, respectively).
 - For example, respondents who reported that they received inpatient treatment and reported lifetime use of marijuana and cocaine but never used alcohol were assigned a value of SUTINDG=17.
 - This edit was implemented for the variables SUTINODG or SUTOUTODG because the specific drug (or drugs) for which respondents received treatment could not be determined. However, it could be determined that respondents received treatment for their use of drugs but not alcohol.
- Respondents who reported lifetime use of both alcohol and drugs were assigned a code of 19 (Substance not specified) to the edited variable corresponding to inpatient or outpatient treatment for other drug use.
 - For example, respondents who reported that they received inpatient treatment and reported lifetime use of alcohol and marijuana were assigned a value of SUTINDG=19.
 - As for the previous scenario, this edit was implemented for the variables SUTINODG or SUTOUTODG because the specific substance (or substances) for which respondents received treatment could not be determined.

8.4.2.1.2 *Editing for Respondents Who Did Not Receive Inpatient or Outpatient Treatment*

Questions about the receipt of inpatient or outpatient treatment for the use of specific substances did not apply to respondents who did not report that they received treatment in inpatient or outpatient locations. Editing differed according to whether respondents answered “no” to all questions for inpatient or outpatient locations, or whether they had missing data for some or all locations, with no reports that they received treatment in the past 12 months.

The following edits to variables for the receipt of treatment for the use of specific substances were implemented if respondents answered “no” to all questions for inpatient or outpatient locations:

- Variables corresponding to substances that respondents never used were assigned codes of 93, as described in Section 8.4.2.1.1. Even if respondents had reported inpatient or outpatient treatment in the past 12 months, the questions for treatment for the use of these substances did not apply.
- Variables corresponding to substances that respondents used in their lifetime were assigned codes of 99 (LEGITIMATE SKIP). These questions did not apply because respondents did not receive inpatient or outpatient treatment.
- If it was unknown whether respondents used a substance in their lifetime, variables corresponding to these substances also were assigned codes of 99. These questions did not apply, even though respondents’ status as lifetime users of specific substances was unknown.
- The variables for the receipt of treatment for the use of some other drug (SUTINODG and SUTOUTODG for inpatient and outpatient treatment, respectively) also were assigned codes of 99 because respondents did not receive treatment in the relevant locations.

The following edits to variables for the receipt of treatment for the use of specific substances were implemented if respondents had missing data for some or all locations, with no reports that they received treatment in the past 12 months:

- Variables corresponding to substances that respondents never used were assigned codes of 93, as described in Section 8.4.2.1.1. Even if respondents had answered “yes” to some questions for inpatient or outpatient treatment in the past 12 months, the questions for treatment for the use of these substances did not apply.
- Variables corresponding to substances that respondents used in their lifetime retained codes of 98 (BLANK) to indicate that the questions would have applied if respondents had reported that they received inpatient or outpatient treatment.
- If it was unknown whether respondents used a substance in their lifetime, variables corresponding to these substances also retained codes of 98. For these respondents, information was unknown for both their receipt of inpatient or outpatient treatment in the past 12 months and whether they had used a substance in their lifetime.
- The variables for the receipt of treatment for the use of some other drug (SUTINODG and SUTOUTODG) retained codes of 98.

8.4.2.2 Editing of Variables for Treatment for Specific Substances Received via Telehealth or in a Prison, Jail, or Juvenile Detention Center

Respondents who reported treatment via telehealth or in a prison, jail or juvenile detention center in the past 12 months were asked a follow-up question for whether the treatment was for alcohol use only, for drug use only, or for both alcohol and drug use. If respondents reported that they did not receive treatment via telehealth or in a prison, jail, or juvenile detention center, then the corresponding edited variables SUTPHVIADB (treatment via telehealth) and SUTPRISADB (treatment in a prison, jail, or juvenile detention center) were assigned legitimate skip codes. If respondents did not know or refused to report whether they had received treatment for these questions, then SUTPHVIADB and SUTPRISADB retained codes of 98 (BLANK).

Respondents who had previously reported lifetime use of alcohol only or lifetime use of drugs only and reported treatment via telehealth or in a prison, jail, or juvenile detention center were skipped out of the follow-up questions.¹⁴⁶ The following edits were implemented when this pattern occurred, and there was no unknown information for the lifetime use of alcohol or drugs:

- If respondents reported lifetime use of alcohol only, then a code of 4 (For alcohol use only LOGICALLY ASSIGNED) was assigned to the corresponding edited variables SUTPHVIADB (treatment via telehealth) and SUTPRISADB (treatment in a prison, jail, or juvenile detention center).
- If respondents reported lifetime use of drugs only, then a code of 5 (For drug use only LOGICALLY ASSIGNED) was assigned to the corresponding edited variables.

If respondents reported lifetime use of alcohol only, but there was unknown information on whether they used drugs in their lifetime, then the edited variables SUTPHVIADB and SUTPRISADB corresponding to the skipped questions retained values of 98 (BLANK). These variables also retained values of 98 if respondents reported lifetime use of drugs, but there was unknown information on whether respondents ever used alcohol.

[Table B.17](#) discusses additional edits for these variables in more detail. For example, [Table B.17](#) discusses edits when respondents were logically inferred to have received treatment via telehealth or in a prison, jail, or juvenile detention center based on their answers to the "OTHER, Specify" questions for the receipt of inpatient or outpatient treatment in some other place.

8.4.3 Other Alcohol or Drug Use Services

As noted previously, respondents also were asked whether they received other services in the past 12 months for their alcohol or drug use, including (1) participation in a support group, (2) help from a peer support specialist or recovery coach, (3) being seen in an emergency room or emergency department, and (4) receipt of detoxification services to manage symptoms of

¹⁴⁶ The follow-up questions asked "When you received the following treatment services in the past 12 months, was it for your alcohol use only, your drug use only, or both your alcohol and drug use?"

withdrawal from alcohol or drug use. These other services were not classified as “substance use treatment.”

8.4.3.1 Editing of Variables for the Receipt of Other Alcohol or Drug Use Services

Respondents could specify in questions TXSBIPDO or TXSBOPGO that the other location where they received “inpatient” or “outpatient” mental health treatment was in a support group, from a peer support specialist or recovery coach, in an emergency room or emergency department, or as detoxification services to manage symptoms of withdrawal from alcohol or drug use. In these situations, the edited variables SUTSUPPGRP (support group), SUTPEERSPC (peer support specialist or recovery coach), SUTERED (emergency room or department), and SUTDETOX (detoxification services) were assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP or SUTOUTOPLSP)].

In addition, respondents were routed to the question regarding the receipt of detoxification services in the past 12 months only if they had previously reported use of alcohol or drugs in the past 12 months.¹⁴⁷ Therefore, respondents who reported lifetime use of alcohol or drugs but were not asked about detoxification services because they did not meet the criteria for the use of alcohol or drugs in the past 12 months were assigned a code of 93 (DID NOT USE ALCOHOL OR DRUGS IN PAST 12 MOS) to the edited variable SUTDETOX.

8.4.3.2 Editing of Variables for the Substances for Which Respondents Received Other Alcohol or Drug Use Services

Respondents who reported that they participated in a support group, received help from a peer support specialist or recovery coach, were seen in an emergency room or emergency department, or received detoxification services were asked a follow-up question for whether the services were for alcohol use only, for drug use only, or for both alcohol and drug use. The principles that were described in Section 8.4.2.2 also were applied to the editing of variables pertaining to the substances for which respondents received other services. For example, if respondents reported lifetime use of alcohol only, then a code of 4 (For alcohol use only LOGICALLY ASSIGNED) was assigned to the corresponding edited variables SUTSUPPADB (for support groups), SUTPEERADB (for peer support specialists or recovery coaches), SUTEREDADB (for emergency rooms or departments), and SUTDETXADB (for detoxification). Similarly, if respondents reported lifetime use of drugs only, then a code of 5 (For drug use only LOGICALLY ASSIGNED) was assigned to the corresponding edited variables.

[Table B.17](#) discusses additional edits for these variables in more detail. For example, [Table B.17](#) discusses edits when respondents were logically inferred to have received these other services based on their answers to the “OTHER, Specify” questions for the receipt of inpatient or outpatient treatment in some other place.

¹⁴⁷ An additional requirement for alcohol and marijuana was that respondents must have reported use of these substances on more than 5 days in the past 12 months.

8.4.4 Perceived Unmet Need for Alcohol and Drug Use Treatment

NSDUH respondents could have perceived an unmet need for substance use treatment if they did not report that they received inpatient substance use treatment; outpatient substance use treatment; treatment via telehealth; treatment in a prison, jail, or juvenile detention center; or MAT for alcohol or drug (i.e., opioid) use in the past 12 months. For these respondents, the perceived unmet need for substance use treatment was measured through the following sets of questions:

- If respondents also did not report that they received services in a support group, from a peer support specialist or recovery coach, in an emergency room or emergency department, or from detoxification services, they were asked one or both of the following questions:
 - These respondents first were asked whether they sought professional counseling, medication, or other treatment for their alcohol or drug use (question TXSBSEK).
 - Respondents who did not report in TXSBSEK that they sought counseling, medication, or other treatment were asked whether they thought they should get professional counseling, medication, or other treatment for their alcohol or drug use (question TXSBGET).
- If respondents reported that they received services in a support group, from a peer support specialist or recovery coach, in an emergency room or emergency department, or from detoxification services, they were asked one or both of the following questions:
 - These respondents first were asked whether they sought additional professional counseling, medication, or other treatment for their alcohol or drug use (question TXSBSEK2).
 - Respondents who did not report in TXSBSEK2 that they sought additional counseling, medication, or other treatment were asked whether they thought they should get additional professional counseling, medication, or other treatment for their alcohol or drug use (question TXSBGET2).

An important aspect of processing the variables for the perceived unmet need for substance use treatment involved assignment of legitimate skip codes, where relevant (Section 2.4.2), according to how respondents answered prior questions in the alcohol and drug use treatment section. For example, if respondents reported that they received inpatient treatment, outpatient treatment, treatment via telehealth, treatment in a prison, jail, or juvenile detention center, or MAT, then all variables pertaining to the perceived unmet need for substance use treatment were assigned legitimate skip codes.

[Table B.17](#) discusses specific edits pertaining to the variables for whether respondents sought treatment or thought they should get treatment for their use of alcohol or drugs.

8.4.5 Reasons for Not Receiving Alcohol or Drug Use Treatment

Respondents who reported an unmet need for substance use treatment in the past 12 months (i.e., they sought or thought they should get substance use treatment or additional substance use treatment) were asked a series of questions regarding reasons they did not receive treatment in that period. For each reason, respondents were asked whether that was one of the reasons or not one of the reasons they did not get professional counseling, medication, or other treatment for their alcohol or drug use (or they did not get additional professional counseling, medication, or other treatment for their alcohol or drug use).

Respondents were assigned a legitimate skip code of 99 to these edited variables in the following situations:

- They reported receiving inpatient or outpatient treatment, treatment via telehealth, treatment in a prison, jail, or juvenile detention center, or MAT.
- They were asked whether they sought treatment or thought they should get treatment (or additional treatment) for their substance use, but they reported that they did not seek treatment and they did not think that they should get it.

8.4.6 Imputation

Missing data for variables related to substance use treatment were imputed using the modified predictive mean neighborhood (modPMN) methodology, as described in Chapter 3. Section 8.4.6.1 describes how these variables were segregated into imputation sets. Section 8.4.6.2 describes how the variables were ordered within the imputation sets. Section 8.4.6.3 provides details on the steps of modPMN and the sequencing of steps.

8.4.6.1 Imputation Sets

The substance use treatment variables were divided into two imputation sets as shown in [Tables 8.1](#) and [8.2](#). These imputation sets are independent of each other and were run in parallel. As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rates. [Tables 8.1](#) and [8.2](#) list the final imputation-revised substance use treatment variables in each imputation set, along with the order in which they were imputed. An imputation indicator was delivered with each imputation-revised variable. Similar to the naming convention of other imputation indicators, the prefix IR of the imputation-revised variable is replaced with II.

Table 8.1 Substance Use Treatment Variables in Imputation Set 1

Variable Description	Final Imputed Variable	Imputation Order
Received Inpatient Substance Use Treatment		
Hospital	IRSUTINHOSP	11
Residential Drug or Alcohol Rehab or Treatment Center	IRSUTINRHAB	5
Residential Mental Health Treatment Center	IRSUTINRSMH	3
Some Other Place as an Inpatient	IRSUTINOPL	4
Received Outpatient Substance Use Treatment		
Drug or Alcohol Rehab or Treatment Center	IRSUTOUTRHAB	10
Mental Health Treatment Center	IRSUTOUTMHCR	12
Office of a Therapist, Psychologist, Psychiatrist, or Substance Use Treatment Professional	IRSUTOUTTHRP	8
General Medical Clinic or Doctor's Office	IRSUTOUTDOC	7
Hospital	IRSUTOUTHOSP	13
School Health or Counseling Center	IRSUTOUTSCHL	6
Some Other Place as an Outpatient	IRSUTOUTOPL	14
Received Other Treatment or Services for Substance Use		
Medication-Assisted Treatment for Alcohol Use	IRSUTRXALC	1
Medication-Assisted Treatment for Opioid Use	IRSUTRXDRG	15
Telehealth Treatment	IRSUTPHVID	2
Prison, Jail, or Juvenile Detention Center	IRSUTPRISON	9

Table 8.2 Substance Use Treatment Variables in Imputation Set 2

Variable Description	Final Imputed Variable	Imputation Order
Received Other Treatment or Services for Substance Use		
Support Group	IRSUTSUPGRP	2
Emergency Room/Emergency Department	IRSUTERED	1
Detoxification/Withdrawal Support Services	IRSUTDETOX	4
Peer Support Specialist or Recovery Coach	IRSUTPEERSPC	3

8.4.6.2 Imputation Process

The substance use treatment variables are all dichotomous. Imputation of categorical variables using the modPMN method is described in Sections 3.6.1 through 3.6.3.

8.4.6.2.1 *Response Propensity Step*

The overall process for the response propensity step is described in Section 3.6.1. For the substance use treatment variables, specifically, two imputation cycles were run for each set. This required fitting 30 response propensity models for set 1 and 8 models for set 2. The following are the four key inputs to the response propensity adjustment step:

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain for most of the models included lifetime drug or alcohol users. There were three exceptions to this: The domain for medication-assisted treatment for alcohol use was lifetime alcohol users, the domain for medication-assisted treatment for opioid use was lifetime users of heroin or prescription pain relievers, and the domain for detoxification support services was past year users of drugs or alcohol.
- **Item Response Indicator:** If a valid response was provided for the given outcome, the individual was deemed an item respondent.
- **Covariates:** Appendix D provides information on the covariates used in each of the response propensity models.

8.4.6.2.2 *Prediction Modeling Step*

The SUDAAN® LOGISTIC procedure (called RLOGIST in SAS®-callable SUDAAN) was used to model dichotomous variables like the substance use treatment variables.¹⁴⁸ All substance use treatment variables had the same levels for the outcome variable:

- received the specified treatment or service in the past 12 months, or
- did not receive the specified treatment or service in the past 12 months.

8.4.6.2.3 *Collocated Stochastic Imputation Step*

This section provides specific information about the collocated stochastic imputation step for the substance use treatment variables. Section 3.6.3 provides general information about the collocated stochastic imputation step. Each of the prediction models is followed by an imputation step where the predicted probabilities from the previous step, along with a random number, is used to impute whether the item nonrespondent received the specified treatment or service. Partial information is not available for these edited variables. Therefore, no adjustments to the predictive mean vectors are made.

8.5 Consumption of Alcohol

The consumption of alcohol section covered special topics related to respondents' use of alcohol. Administration of alcohol-related questions focused on respondents who had ever used alcohol, and most questions were limited to respondents who reported alcohol use in the past 30 days.

¹⁴⁸ SUDAAN® is a registered trademark of Research Triangle Institute. SAS® software is a registered trademark of SAS Institute Inc.

Respondents were asked questions about the following topics in the consumption of alcohol section:

- number of drinks consumed on the last occasion of alcohol use in the past 30 days;
- the last occasion of alcohol use in the past 30 days by adolescents or young adults younger than the legal drinking age of 21, including who respondents were with, where they drank, and how they obtained the alcohol;
- use of other drugs in combination with alcohol (i.e., at the same time or within a couple of hours of using alcohol) in the past 30 days; and
- history of binge alcohol use (i.e., consumption of five or more drinks on a single occasion for males or four or more drinks on a single occasion for females).

The remainder of this section documents the editing procedures.

8.5.1 General Skip Logic Edits for Alcohol-Related Variables

One of the important aspects of processing variables in the consumption of alcohol section consisted of assigning codes of 91, 93, and 99 to variables that had been skipped because the questions did not apply. Certain questions may not apply to respondents for multiple reasons, such as if they never used alcohol or if they used alcohol in the past month but they were aged 21 or older.

[Table B.18](#) discusses the prioritization of skip logic edits for alcohol-related variables in this section. For example, if respondents had never used alcohol, variables in the consumption of alcohol section were assigned a code of 91, 991, or 9991 (NEVER USED ALCOHOL), regardless of whether questions were not applicable for other reasons (e.g., if respondents were aged 21 or older). This edit kept data for a constant number of respondents coded as indicating nonuse of alcohol for most alcohol-related variables in the section.

8.5.2 Number of Drinks for the Last Alcohol Use in the Past 30 Days

Other than the implementation of edits based on the skip logic for alcohol-related variables (see [Table B.18](#)), only limited editing was done for the variables pertaining to the number of drinks that respondents consumed the last time they drank in the past 30 days (question CA01; edited variable CADRLAST). However, data from CADRLAST were used in editing the binge alcohol history variables (Section 8.5.5).

8.5.3 Source and Location of Last Underage Alcohol Use in the Past 30 Days

Other than implementation of edits based on the skip logic for alcohol-related variables (see [Table B.18](#)), only limited editing was done for the variables pertaining to alcohol use in the past 30 days among adolescents or young adults aged 12 to 20 (questions CA02A through CA07SP). In question CA07 (edited variable CAFREWHO), for example, respondents were asked to give one response for how they got the last alcoholic beverage that they drank, including the option to report that they got it “some other way.” Respondents who reported that they got the alcohol some other way were asked to specify what this other way was (edited variable

CAFRESP). If a response was given in CAFRESP (corresponding to CA07SP) that corresponded to one of the existing response options in question CA07, then CAFREWHO was assigned the relevant response plus a value of 10. For example, if CAFRESP indicated that respondents got their last alcoholic beverage from a nonrelative who was 21 or older who gave the alcohol to the respondent (i.e., equivalent to CA07 = 3), then CAFREWHO was assigned a code of 13.

No editing was done to CABUYFRE (corresponding to question CA03), CAGVMONY (corresponding to question CA04), CABUYWHO (corresponding to question CA05), or CABPLACE (corresponding to question CA05A) if a response to CAFRESP was inconsistent with a response to an earlier question. The rationale for not editing these variables was that insufficient information for editing these variables might have been available from what respondents specified.

For example, respondents could report in question CA04 that they gave money to someone else who bought their last alcoholic beverage for the respondents, but then report in CAFRESP that the “other” way that they got the last alcoholic beverage they drank was by purchasing it, such as with fake identification. In this example, question CA04 about buying alcohol was asked only if respondents reported in question CA03 that *they* paid for the last alcohol that they drank. Even if the respondents specified buying alcohol in CAFRESP, however, it would not be known whether they bought the alcohol with their own money or whether the alcohol was purchased with someone else’s money.

Questions CA02B (where underage respondents drank alcohol the last time) and CA09 (the specific drugs that respondents of any age used at the same time or within a couple of hours of when they last drank alcohol) were “enter all that apply” questions. As in other sections, the responses were captured as discrete variables. The individual edited variables for these “enter all that apply” variables were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (Section 2.4.4).

For the edited variables corresponding to question CA02B, if respondents specified some other place where they drank alcohol the last time that corresponded to one of the existing response options in question CA02B, and they had not chosen that response, then the corresponding edited variable was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED. For example, if respondents reported that they drank alcohol in someone else’s home the last time they drank but they had not chosen this response in question CA02B, then the edited variable CADROTHM was assigned a code of 3. In addition, codes of 94 and 97 (for “don’t know” and “refused,” respectively) were assigned to the entire list of variables corresponding to question CA02B if respondents did not know or refused to report where they were the last time they drank alcohol.

8.5.4 Use of Illicit Drugs with Alcohol in the Past 30 Days

If respondents drank alcohol in the past 30 days and they also used marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or methamphetamine in that period, they were asked in question CA08 to report whether they used other drugs the last time they drank alcohol in the past 30 days. Respondents who answered CA08 as “yes” and used more than one

of these other drugs in the past 30 days were asked in question CA09 to report the specific other drugs they used the last time they drank alcohol.

CA09 was skipped if respondents answered CA08 as “yes” but reported using only one of these other drugs in the past 30 days. In this situation, it could be logically inferred that the answer of “yes” in CA08 applied to the use of that specific other drug with alcohol. For example, if respondents used marijuana in the past 30 days and did not use cocaine through methamphetamine in the past 30 days (i.e., last used these drugs more than 30 days ago or never used them),¹⁴⁹ then the edited variable CADRKMARJ for the use of marijuana in the past 30 days was assigned a code of 5 (Response entered LOGICALLY ASSIGNED [from skip pattern]). The remaining variables CADRKCOCN through CADRKMETH were assigned a code of 99 (LEGITIMATE SKIP).

For the edited variables CADRKMARJ through CADRKMETH corresponding to question CA09, codes of 94 or 97 were retained only for those drugs that respondents used in the past 30 days or could have used in the past month. Otherwise, codes of 94 or 97 in the individual drug variables corresponding to question CA09 were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned) if respondents used some drug in the past 30 days but not the drug that corresponded to the variable of interest. Suppose, for example, that respondents used marijuana and inhalants in the past 30 days and did not use any other illicit drugs in that period. If question CA08 was answered as “yes” and CA09 was refused, then the edited variables CADRKMARJ and CADRKINHL retained a code of 97 and the remaining drug variables were assigned a code of 89.

Special consideration also was given to the editing of the variables CADRKCOCN, CADRKHERN, and CADRKMETH for the use of cocaine, heroin, or methamphetamine, respectively, during or within a couple of hours of the episode of alcohol use in the past 30 days, because respondents could indicate use of these substances in the past 30 days in an interview section specifically for that substance (e.g., heroin) or in another section (e.g., the special drugs section). If respondents answered question CA08 as “yes,” reported using cocaine, heroin, or methamphetamine in the past 30 days in the special drugs section, and did not use other drugs in the past 30 days, then the edits applied that were described previously when CA08 was answered as “yes” and respondents reported the use of only one of these drugs in the past 30 days. For example, if respondents reported in the special drugs section that they used methamphetamine with a needle in the past 30 days and they did not use marijuana through inhalants in that period, then CADRKMETH was assigned a code of 5, and the other drug variables from question CA09 were coded as 99.

Respondents who answered question CA08 as “yes” and reported the use of more than one drug in the past 30 days would have the response option “Cocaine or ‘crack’” (captured in CADRKCOCN) available to them if they previously reported past month use of cocaine in any of the following places: (1) in the section for any cocaine use, (2) in the section for crack cocaine use, or (3) in the special drugs section (i.e., use of cocaine with a needle in the past 30 days).

¹⁴⁹ If a drug’s recency variable had been coded as 8, the respondent used the drug at some point in the past 12 months, which could include use within the past 30 days.

Similarly, respondents who answered question CA08 as “yes” and reported using more than one drug in the past 30 days could choose the response option “Heroin” (captured in CADRKHERN) if they reported past month use of heroin in the heroin section or if they reported past month sniffing, smoking, or injection of heroin in the special drugs section. For methamphetamine, the response option “Methamphetamine” (captured in CADRKMETH) was available to respondents who indicated past month methamphetamine use in the methamphetamine section or in the special drugs section (i.e., use of methamphetamine with a needle in the past 30 days).

Consequently, respondents who used more than one drug other than alcohol in the past 30 days would have the response options for CADRKCOCN, CADRKHERN, or CADRKMETH available to them if they reported past month use of the relevant drug(s) in the special drugs section, even if they had not reported past month use in the corresponding earlier section(s) of the interview for those specific drugs. In most situations, no further editing was done to these data, even though the corresponding earlier substance use data did not indicate use in the past 30 days.

8.5.5 Binge Alcohol History

The threshold for binge alcohol use in the past 30 days for females is the consumption of four or more drinks on an occasion, and the threshold for males is the consumption of five or more drinks on an occasion.¹⁵⁰ In turn, the binge alcohol history questions CA10 through CA11D ask females about their history of consumption of four or more drinks on a single occasion. For males, questions CA10 through CA11D ask about the history of consumption of five or more drinks on an occasion.

If respondents were lifetime alcohol users, and other data from the alcohol section or from CADRLAST did not indicate that respondents had been binge drinkers in the past 30 days,¹⁵¹ then they were asked in question CA10 (edited variable CABINGEVR) whether they had ever engaged in binge drinking in their lifetime. Those respondents who had ever been binge drinkers (or had reported binge drinking in the past 30 days in the alcohol section or in CADRLAST) were asked in question CA11 (edited variable CABINGAGE) to report how old they were the first time they engaged in binge drinking. If CABINGAGE was within 1 year of respondents’ current age, they were routed to questions CA11A through CA11D (edited variables CABINGYFU and CABINGMFU) to report the year and month when they first engaged in binge drinking.

As noted previously, the routing logic for these questions took into account respondents’ answers in the earlier alcohol section. Specifically, the CAI program created a flag variable called BINGEFLAG. The default value for this flag was 0. The flag was coded as 1 if question

¹⁵⁰ The threshold for females changed in 2015 from consumption of five or more drinks on an occasion to the consumption of four or more drinks on an occasion.

¹⁵¹ For brevity, the remaining text in this section refers to reports of binge alcohol use or binge drinking, defined as consumption of four or more drinks on a single occasion on at least 1 day in the past 30 days for females or as consumption of five or more drinks on a single occasion on at least 1 day in this period for males.

AL08 in the alcohol section indicated that respondents had been binge drinkers in the past 30 days. Also, if question AL07 indicated that respondents usually had four or more drinks (for females) or five or more drinks (for males) on those days when they drank in the past 30 days, then this flag was coded as 1.

The edited variable in the consumption of alcohol section corresponding to this flag was named CABINGFLG. If the variables ALCUS30D (corresponding to question AL07) and ALCBNG30D (corresponding to question AL08) indicated binge drinking in the past 30 days but CABINGFLG did not already have a value of 1, then CABINGFLG was set to a logically assigned value of 11. If ALCUS30D and ALCBNG30D had been set to bad data as a result of edits in the alcohol section (see Section 6.2.4.7 and [Table B.5](#)), then CABINGFLG also was set to bad data. In turn, any questions that had been answered in CA11 through CA11D were set to bad data.

[Table B.19](#) discusses additional edits associated with the binge alcohol history variables in the consumption of alcohol section. The first set of edits pertains to the consistency checks between the age when respondents reported that they first engaged in binge drinking and the age at initiation of binge drinking based on the year and month when respondents reported that they first engaged in this behavior.

[Table B.19](#) then presents edits for these binge alcohol variables relative to respondents' ages and to corresponding alcohol use data from earlier in the interview. For example, if respondents reported in question CA11 that they first engaged in binge drinking at an age that was earlier than what they reported for when they first had a drink of any alcohol (from the alcohol use variable ALCTRY), then CABINGAGE was set to bad data. Any nonmissing values in CABINGYFU and CABINGMFU also were replaced with codes for bad data. Similarly, suppose respondents first used any alcohol at their current age. If they had a lifetime history of binge drinking (including any indication of this behavior in the past 30 days) but CA11 had a missing value (e.g., "don't know" or "refused"), then it could be logically inferred in CABINGAGE that the first episode of binge alcohol use occurred at the respondents' current age.

In addition, "editing indicator" (EI) variables were created to indicate when nonmissing values were logically assigned to the binge alcohol variables in the consumption of alcohol section. The default value for these EI variables was 1 (i.e., questionnaire data). The EI variables were assigned a value of 2 (Logically assigned data) if a logical inference was made that involved assignment of a nonmissing value to the binge alcohol history variables.

In keeping with the general principle of not editing across sections of the interview (Section 2.4.1), no editing was done to the data in the earlier alcohol section when a respondent's reported first binge alcohol use suggested more recent use of alcohol than was indicated in the recency variable ALCREC. For example, ALCREC may not indicate use in the past month or past year, but respondents could report that they first had five or more drinks for males or four or more drinks for females on a single occasion at their current age. Although these data from the consumption of alcohol would suggest that these respondents used alcohol in the past 12 months, no editing was done to ALCREC based on these data.

8.6 Emerging Issues

Variables were edited in 2023 for the following measures in the emerging issues section:¹⁵²

- respondents' perceived recovery from substance use or mental health problems;
- use of kratom;
- vaping of only flavoring;
- use of synthetic marijuana;
- use of synthetic stimulants; and
- use of IMF.

As noted in Section 2.3.2.2, responses in the emerging issues section were replaced with bad data codes when patterned responses were identified in this section.

8.6.1 Perceived Recovery from Substance Use or Mental Health Problems

Adult respondents were asked in question CA15 whether they thought they ever had a problem with their drug or alcohol use, regardless of whether they previously reported any use of drugs or alcohol. Similarly, all adults were asked in question CA17 whether they ever had a problem with their mental health. Respondents were asked follow-up questions about whether they considered themselves to be in recovery (or recovered) if they answered the corresponding lead question as "yes." For example, adults who reported in CA15 that they ever had a problem with their drug or alcohol use were asked in question CA16 whether they considered themselves to be in recovery or to have recovered from their drug or alcohol use.

An important aspect of the processing of variables pertaining to perceived recovery consisted of assigning codes of 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. All adolescent respondents aged 12 to 17 were assigned a code of 99 for the edited recovery variables. In addition, variables for perceived recovery from a substance use or mental health problem were assigned a code of 99 if adults answered the corresponding lead question as "no." If adults did not know or refused to report whether they ever had a problem with their substance use or mental health, then the follow-up variables for perceived recovery (e.g., perceived recovery from a substance use problem) retained codes of "blank."

Consistent with the general principle of not editing across sections of the interview (Section 2.4.1), data from these perceived recovery items were not edited for consistency with data from other sections, or vice versa. For example, if adult respondents reported that they never had a problem with their mental health, these data were not edited for consistency with reports of symptoms of major depressive episode in the adult depression section (Chapter 10). Similarly, data on perceptions of substance use problems or recovery from substance use problems were not edited for consistency with earlier data on substance use (Chapter 6), SUDs (Section 7.3), or the receipt

¹⁵² Questions in the emerging issues section regarding vaping of marijuana continued to be asked in 2023. However, edited and imputed variables corresponding to these questions were not created because questions were added to the marijuana section in 2022 to capture information regarding ways that marijuana was used, including use by vaping. These new questions were intended to replace questions asked in the emerging issues section about vaping of marijuana.

of substance use treatment (Section 8.4). Consequently, NSDUH data users could have different impressions about adults' substance use problems, mental health issues, or recovery, depending on whether data users analyze data from these perceived recovery items or from other sections.

8.6.2 Kratom

Respondents were asked in the emerging issues section about lifetime use and most recent use of kratom. Minimal editing was done to create the variables KRATEVER and KRATREC.

Respondents who reported never using kratom (KRATEVER = 2) were assigned a value of 91 (NEVER USED KRATOM) to the edited variable KRATREC. If respondents reported lifetime use of kratom (KRATEVER = 1), but they answered KALAST3 as DK or REF, then KRATREC was assigned a code of 9, where 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

8.6.3 Vaping of Only Flavoring

Respondents were asked in the emerging issues section about lifetime and most recent vaping of only flavoring. Minimal editing was done to create the variables FLVVAPEVR and FLVVAPREC. Respondents who reported never vaping of only flavoring (FLVVAPEVR = 2) were assigned a value of 91 (NEVER VAPED ONLY FLAVORING) to the edited variable FLVVAPREC. If respondents reported lifetime vaping of only flavoring (FLVVAPEVR = 1), but they answered question VPFLREC as DK or REF, then FLVVAPREC was assigned a code of 9, where 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

8.6.4 Synthetic Marijuana and Synthetic Stimulants

Respondents were asked questions SMLIF, SMREC, SSLIF, and SSREC about lifetime use and most recent use of synthetic marijuana and synthetic stimulants. The edited variables for these questions were SYNMJRJEVR and SYNMJRJREC for lifetime use and most recent use of synthetic marijuana, respectively, and SYNSTMEVR and SYNSTMREC for lifetime use and most recent use of synthetic stimulants, respectively.

Minimal editing was done to these variables. Respondents who reported never using synthetic marijuana (SYNMJRJEVR = 2) were assigned a value of 91 (NEVER USED SYNTHETIC MARIJUANA) to the edited variable SYNMJRJREC. Similarly, respondents who reported never using synthetic stimulants (SYNSTMEVR = 2) were assigned a value of 91 (NEVER USED SYNTHETIC STIMULANTS) to the edited variable SYNSTMREC. If respondents reported lifetime use of synthetic marijuana or stimulants, but they answered the corresponding follow-up recency question as "don't know" or "refused," then the respective edited recency variable was assigned a code of 9, where 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

8.6.5 Illegally Made Fentanyl

New questions were added in 2022 regarding lifetime use and most recent use of IMF. IMF was defined for respondents as fentanyl that people cannot get from a doctor or pharmacy and that can come in forms such as powder, pills, or blotter paper. Minimal editing was done to the edited variables IMFEVER and IMFREC. Respondents who reported never used IMF (IMFEVER = 2) were assigned a value of 91 (NEVER USED ILLEGALLY MADE FENTANYL) to the edited

variable IMFREC. If respondents reported lifetime use of IMF (IMFEVER = 1), but they answered question IFREC as DK or REF, then IMFREC was assigned a code of 9, where 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

8.6.6 Imputation

Missing data for variables related to IMF, kratom, synthetic marijuana, synthetic stimulants, and vaping of flavoring were imputed using the modPMN methodology, as described in Chapter 3. Section 8.6.6.1 describes how these variables were segregated into imputation sets. Section 8.6.6.2 describes how the variables were ordered within the imputation sets. Section 8.6.6.3 provides details on the steps of modPMN and the sequencing of steps.

8.6.6.1 Imputation Sets

The emerging issues variables imputed using the modPMN methodology are segregated into four imputation sets as outlined in [Table 8.3](#). Set 1 can be run in parallel with set 2. Likewise, set 3 can be run in parallel with set 4.

Table 8.3 Substances and Substance Use Measures

Substance	Substance Use Measures	
	Lifetime Usage	Recency of Use
Illegally Made Fentanyl	Set 1	Set 3
Kratom		
Synthetic Marijuana		
Synthetic Stimulants		
Vaping of Flavoring	Set 2	Set 4

8.6.6.2 Imputation Order within Imputation Sets

As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rates. The sequencing for 2023 is shown in [Table 8.4](#). Sets 2 and 4 each contained only one substance, vaping of flavoring, and therefore are not included in [Table 8.4](#).

Table 8.4 Sequencing of Substances within Imputation Sets, 2023

Imputation Set	Sequence
Lifetime Usage: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	Synthetic Stimulants, Illegally Made Fentanyl, Synthetic Marijuana, Kratom
Recency of Use: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	Synthetic Marijuana, Kratom, Illegally Made Fentanyl, Synthetic Stimulants

8.6.6.3 Imputation Process

The imputation procedures used to impute IMF, kratom, synthetic marijuana, synthetic stimulants, and vaping of flavoring followed the steps of modPMN for categorical variables, as described in Sections 3.6.1 through 3.6.3. Additional details specific to these variables are included below.

8.6.6.3.1 Response Propensity Step

General information about the response propensity adjustment step of modPMN is provided in Section 3.6.1. The total number of response propensity models fit for each of these imputation sets can be found in [Table 8.5](#). Following are the four key inputs to the response propensity adjustment step:

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain for each set is detailed in [Table 8.5](#).
- **Item Response Indicator:** If a valid response was provided for the given measure for the substance, the individual was deemed an item respondent.
- **Covariates:** Appendix D provides information on the covariates used in each of the response propensity models.

Table 8.5 Summary of Response Propensity Models

Set	Number of Response Propensity Models	Domain
1: Lifetime Usage: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	8	All respondents
2: Lifetime Usage: Vaping of Flavoring	1	All respondents
3: Recency of Use: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	8	Lifetime users, based on imputation-revised lifetime use
4: Recency of Use: Vaping of Flavoring	1	Lifetime users, based on imputation-revised lifetime use

8.6.6.3.2 Prediction Modeling Step

The SUDAAN MULTILog procedure was used to model multilevel categorical variables in sets 3 and 4. The SUDAAN LOGISTIC procedure was used to model the dichotomous variables in sets 1 and 2. The levels for outcome variables for each set are listed in [Table 8.6](#).

Table 8.6 Outcomes of Prediction Modeling

Set	Outcome Variable Levels
1: Lifetime Usage: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	1: Lifetime Use 2: Never Used
2: Lifetime Usage: Vaping of Flavoring	1: Lifetime Use 2: Never Used
3: Recency of Use: Illegally Made Fentanyl, Kratom, Synthetic Marijuana, and Synthetic Stimulants	1: Past Month Use 2: Past Year, Not Past Month Use 3: Lifetime, Not Past Year Use
4: Recency of Use: Vaping of Flavoring	1: Past Month Use 2: Past Year, Not Past Month Use 3: Lifetime, Not Past Year Use

8.6.6.3.3 Collocated Stochastic Imputation Step

Imputation of these variables followed the collocated stochastic imputation method, as described in Section 3.6.3. No adjustments to the predictive mean vector are made to sets 1, 2, and 4 because partial information is not available for the edited variables in these sets. Partial information from the edited variables was used to adjust the predictive mean vectors in set 3. For more information on these restrictions, see Appendix E.

8.6.6.3.4 Final Variables

No final imputation-revised variables indicating lifetime usage (sets 1 and 2) alone were created because this information was recoded in the final imputation-revised recency-of-use variables. Imputation indicators also were not created.

The final imputation-revised variables for the three recency-of-use sets followed a similar pattern: the IR prefix followed by a drug abbreviation and ending with REC to indicate recency of use. Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

8.6.6.4 Recodes for Additional Analyses

From the final imputation-revised variables, some additional recodes were created to use in additional analyses. For recency-of-use variables, this included three dichotomous indicator variables to indicate lifetime use (XXXFLAG), past year use (XXXYR), and past month use (XXXMON).

[Table 8.7](#) contains information on all dichotomous indicator recodes created.

Table 8.7 Prevalence Recodes

Drug	Variable Names
Illegally Made Fentanyl	IMFFLAG, IMFYR, IMFMON
Kratom	KRATOMFLAG, KRATOMYR, KRATOMMON
Synthetic Marijuana	SYNMRJFLAG, SYNMRJYR, SYNMRJMON
Synthetic Stimulants	SYNSTMFLAG, SYNSTMYR, SYNSTMFLAG
Vaping of Flavoring	FLVVAPFLAG, FLVVAPYR, FLVVAPMON

8.7 Market Information for Marijuana

The market information for marijuana section underwent considerable revisions for 2022. Consequently, all edited variables were new or renamed for 2022. This section focused on how past year marijuana users acquired marijuana or cannabis products in the past 12 months, the form of marijuana or cannabis products that they purchased, the quantity of joints or loose marijuana that they purchased, and the amount of money spent overall on marijuana, and on marijuana in joints or loose form.

Respondents who reported in the marijuana section of the interview that they used marijuana or cannabis products in the past year were first asked how they obtained marijuana in the past 12 months. Respondents were asked individual yes/no questions for whether they paid for or bought it, traded something else for it, got it for free or shared someone else's, grew it, or got it in some other way.

Edits in this section principally involved assigning appropriate legitimate skip codes based on the logic for determining whether respondents should be asked questions in the section (see Section 2.4.2). If respondents reported in the marijuana section that they never used marijuana, then the edited variables for marijuana purchases were assigned a code of 91, 991, or 9991 (NEVER USED MARIJUANA), depending on the allowable length of the response options. Similarly, if respondents' edited marijuana recency MJREC indicated that they last used marijuana more than 12 months ago, then the edited variables for marijuana purchases were assigned a code of 93, 993, or 9993 (DID NOT USE MARIJUANA IN THE PAST 12 MONTHS). If respondents were not asked the questions in the marijuana purchases section, but their edited marijuana recency had a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), then the skipped marijuana purchases variables retained codes of 98, 998, or 9998 [BLANK (NO ANSWER)]. Some of these respondents could have used marijuana in the past 12 months and would have been eligible to be asked questions in this section.

If respondents previously reported that they used marijuana in the past 12 months, a key aspect of editing of variables for marijuana purchases involved assignment of legitimate skip codes (99, 999, or 9999) according to how respondents were routed through the section. For example, respondents who reported that they did not pay for or buy marijuana in the past 12 months were not asked questions about the form of marijuana they last purchased, the amount of marijuana they purchased, and the amount they paid for marijuana. If respondents reported

in question MJE01A that they paid for or bought marijuana, but they reported that they did not buy marijuana in prerolled joints or loose form in questions MJE05A or MJE05B, respectively, then they were not asked the corresponding questions about the amount they paid for marijuana in these forms.

Respondents who reported in question MJE01E that they got marijuana or cannabis products in some other way in the past 12 months were asked in “OTHER, Specify” question MJE01E2 to describe the other way they got marijuana. Respondents could provide up to five ways in which they got marijuana or cannabis products in the past 12 months. If respondents reported in question MJE01E that they did not get marijuana in some other way, then the edited “OTHER, Specify” variables corresponding to MJE01E2 were assigned legitimate skip codes.

Consistent with the procedures described in Section 2.4.5, if respondents did not report getting marijuana or cannabis products in the past 12 months in a particular way in questions MJE01A to MJE01D, but they specified a particular way in question MJE01E2 as some other way they got marijuana in the past 12 months, then the corresponding edited variable for that particular way was assigned a code of 3 (Yes LOGICALLY ASSIGNED). For example, if respondents did not report paying for or buying marijuana in the past 12 months in question MJE01A, but they specified in question MJE01E2 that they had bought marijuana, then the corresponding edited variable MKMBGHTYR was assigned a code of 3.

The remaining processing of the variables in the marijuana purchases section involved creating summary variables for the price that respondents paid for the last marijuana or cannabis products they bought. For example, respondents who reported buying marijuana in the past 12 months were first asked in the open-ended write-in question MJE04 to report the total amount they paid the last time they purchased marijuana. Respondents who did not know or refused to report the amount were asked a series of follow-up questions that provided respondents with categorical levels as options to choose. Respondents who were administered these follow-up questions were first asked whether they spent less than \$200 or spent \$200 or more. Respondents were then routed to one of two questions with more detailed price categories (e.g., less than \$25, \$25.00 to \$49.99, \$50.00 to \$99.99, \$100 to \$149.99, \$150.00 to \$199.99) to more specifically define the amount they paid. A composite summary cost variable was created to combine information from the three categorical price questions. In addition, a second composite cost variable was created to combine information from the write-in response with the follow-up categorical price information.

Respondents who reported that they last bought marijuana in prerolled joints and in at least one additional form were asked in a write-in question to report a dollar amount that they last spent on prerolled joints. Similarly, respondents who reported that they last bought marijuana in loose form and in at least one other form were asked to report a dollar amount that they last spent on marijuana in loose form. Respondents who reported that they last bought marijuana in prerolled joints only or in loose form only were not asked these follow-up price questions because it could be inferred that the amount respondents reported for total cost of the last marijuana they purchased (MJE04) was the amount they spent on prerolled joints or loose marijuana, respectively.

However, respondents could provide inconsistent responses for the amount of money they spent on marijuana. For example, if respondents reported that they bought both joints and marijuana in loose form the last time, the individual amounts they reported that they spent on joints and in loose form could sum to amounts that were higher or lower than the total amount respondents previously reported that they last spent on marijuana. Rather than edit for these inconsistencies, the indicator variable MKMCOMPAMT was created to alert users to this situation.

9. Editing of Risk and Protective Factors for Substance Use Variables

9.1 Introduction

This chapter discusses editing procedures for the following sections that pertain to risk and protective factors for substance use:

- risk/availability,
- special topics,
- social environment (administered to adults aged 18 or older),
- youth experiences (administered to adolescents aged 12 to 17), and
- parenting experiences (administered to parents when an adolescent aged 12 to 17 also was selected for a separate interview).

These questionnaire sections, which are self-administered for both in-person and web respondents, are included in this chapter according to their related topical content rather than in the sequential order in which they appeared to respondents during the interview. The order in which these sections were presented to respondents relative to other sections of the interview can be found in the computer-assisted interviewing (CAI) specifications for the 2023 National Survey on Drug Use and Health (NSDUH) that are available on the [NSDUH Data Collection](#) web page.

None of the edited variables in this chapter were imputed to remove missing values. Although the special topics section of the interview includes questions about driving under the influence of alcohol or illegal drugs, it was included in this chapter on risk and protective factors because most of the questions in this section pertain to arrests or other involvement with the criminal justice system (e.g., being on probation or parole in the past 12 months).

9.2 Risk/Availability

The risk/availability section asked about the perceived risk of harm associated with use of alcohol or specific illegal drugs, perceived ease of obtaining different illegal drugs, whether respondents were approached by someone in the past 30 days who was trying to sell an illegal drug, and general risk-taking types of behaviors. The latter included questions on the frequency with which respondents got a “kick out of doing things that are a little dangerous,” how often they tried to test themselves “by doing something a little risky,” and their frequency of seatbelt use.

Minimal processing of data was done to variables in this section. Specifically, the unedited variables were assigned final, descriptive variable names to assist data users in understanding content (e.g., RSKCIGPKD corresponding to question RK01A, which asked about the perceived risk of harm associated with smoking one or more packs of cigarettes per day). As noted in Section 2.3.2.2, responses in the risk/availability section that were determined to be patterned

responses were replaced with “bad data” codes. No further editing or processing was done to the variables in this section.

9.3 Special Topics

The special topics section asked about arrests in the respondents’ lifetime and in the past 12 months, including arrests for specific offenses in the past 12 months (not counting minor traffic violations). This section also included questions about respondents being on probation or parole in the past 12 months, operating vehicles under the influence of alcohol or specific other drugs in the past 12 months, and respondents’ knowledge about their states’ marijuana laws.

9.3.1 Editing of Variables for Criminal Justice Involvement

Questions in the special topics section of the interview that pertained to respondents’ involvement with the criminal justice system included questions about arrests in the respondents’ lifetime and in the past 12 months and whether respondents were on probation or parole in the past 12 months. Being arrested and booked for breaking the law was defined for respondents as being taken into custody and processed by the police or someone connected with the courts, even if respondents were then released. Consequently, respondents who reported that they were arrested and booked were not necessarily jailed for an offense, nor were they necessarily convicted of a crime.

If respondents reported that they had never been arrested in their lifetime and they did not report being on probation or parole in the past 12 months (see below), the edited variables pertaining to arrests in the past 12 months were assigned legitimate skip codes (Section 2.4.2). Other standard edits described in Section 2.4.3 pertaining to situations where respondents answered “don’t know” or “refused” to the lifetime arrest question were applied to the past year arrest variables that had been skipped.

Similarly, if respondents reported being arrested in their lifetime but reported being arrested 0 times in the past 12 months, the questions pertaining to arrests for specific offenses in the past 12 months were skipped and the corresponding edited variables were assigned legitimate skip codes (99). Unlike the logic in most other sections that skipped respondents out of follow-up questions if they answered a lead question as “don’t know” or “refused,” respondents who did not know how many times they were arrested in the past 12 months or who refused to answer this question were asked whether they were arrested for specific offenses in the past 12 months.

[Table B.20](#) in Appendix B discusses additional edits for the criminal justice variables in the special topics section. For example, respondents could report that they had never been arrested in their lifetime but could report that they were on probation, parole, or supervised release in the past 12 months. Because someone could not be on probation or parole without first having been arrested for a crime, these respondents were logically inferred to have been arrested in their lifetime. When this situation occurred, the skipped variables pertaining to arrests in the past 12 months retained a value of “blank.”

Respondents who reported in question SP02 that they were arrested and booked for breaking the law at least once in the past 12 months and gave negative answers to every question about specific arrests (including arrests for “some other offense”) were routed to a follow-up question, SP03R, that asked respondents to verify their previous answer from SP02. If respondents did not indicate that their previous answer from question SP02 was correct, they were routed to question SP03S, where they were allowed to update their answer regarding the number of times they were arrested or booked in the past 12 months. If respondents indicated in SP03S that they had been arrested or booked 0 times in the past 12 months, the edited variable NOBOOKYR was assigned a value of 0. Further, when SP03S indicated that respondents had been arrested and booked 0 times in the past 12 months, it was logically inferred that all items pertaining to specific arrests in the past 12 months should have been skipped. Therefore, all of the variables associated with specific arrests in the past 12 months were assigned a code of 89 = LEGITIMATE SKIP Logically assigned (Section 2.4.2).

If respondents indicated in SP03R that their previous answer from SP02 was correct, or if they reported being arrested and booked for at least one offense in SP03S, they were asked to specify at least one offense for which they were arrested and booked in the past 12 months (questions SP03RSP or SP03SSP). If respondents specified a legitimate offense in SP03RSP or SP03SSP after giving negative answers to every question about specific arrests, the “OTHER, Specify” variable BKOTHOFF was updated to incorporate the response from SP03RSP or SP03SSP. A code of 3 (Yes LOGICALLY ASSIGNED) also was assigned to the “some other offense” variable (BKOTH) to indicate that this edit had taken place (Section 2.4.5).

Any information that respondents specified in SP03RSP or SP03SSP also was used to edit variables pertaining to offenses that respondents previously had been asked about. For example, if respondents had answered question SP03A as “no” (i.e., had not been arrested and booked for motor vehicle theft in the past 12 months) but then specified in SP03RSP or SP03SSP that motor vehicle theft was one of the offenses for which they had been arrested, the edited variable BKMVTHFT was assigned a code of 3 (Yes LOGICALLY ASSIGNED).

If respondents reiterated in SP03R or SP03S that they had been arrested and booked for at least one offense in the past 12 months but still did not report a legitimate offense in questions SP03RSP or SP03SSP (including situations in which they answered SP03RSP or SP03SSP as “don’t know” or “refused”), then a code of 5 (Offense unknown) was assigned to BKOTH. For that reason, the response from SP02 or SP03S was retained in NOBOOKYR to indicate that the respondents were arrested in the past 12 months, but it was not possible to determine the specific crime for which they were arrested.

9.3.2 Editing of Variables for Driving under the Influence

Questions SP06A through SP06G asked about driving under the influence of alcohol, marijuana, cocaine (including crack, where applicable), heroin, hallucinogens, inhalants, and methamphetamine, respectively. Respondents who reported in question SP06A that they drove under the influence of alcohol in the past 12 months and also reported in SP06B through SP06G that they drove under the influence of any of the other substances mentioned previously in that same period were asked in question SP06H if they drove under the influence of alcohol only.

Respondents were skipped out of questions pertaining to driving under the influence of alcohol or the specific other substances if they reported in previous sections of the interview that they did not use the relevant substance(s) in the past 12 months. For example, respondents who did not previously report use of marijuana in the past 12 months in the marijuana section were not asked SP06B about driving under the influence of marijuana in that period.

For cocaine, heroin, and methamphetamine, however, respondents who had previously reported that they last used these substances more than 12 months prior to the interview in the respective substance sections but then reported past year use in the special drugs section of the interview (Section 7.2) also were asked the questions about driving under the influence of these substances. Therefore, the editing of the variables for driving under the influence of cocaine, heroin, or methamphetamine considered reports of use within the past 12 months from both the specific substance sections and the special drugs section. Because of the additional logic for how respondents could be asked the questions about driving under the influence of cocaine, heroin, or methamphetamine, edits to the variables for driving under the influence were implemented for two groups: (1) alcohol, marijuana, hallucinogens, or inhalants; and (2) cocaine, heroin, or methamphetamine.

9.3.2.1 Driving under the Influence of Alcohol, Marijuana, Hallucinogens, or Inhalants

The edits described below were implemented for the variables for driving under the influence of alcohol, marijuana, hallucinogens, or inhalants when respondents were determined to have skipped out of the corresponding questions because of nonuse in their lifetime or in the past 12 months. Edits for the variable for driving under the influence of alcohol only (i.e., but not any of the specific illicit drugs) are discussed separately in Section 9.3.2.3.

- If respondents were lifetime nonusers of the relevant substance, then the corresponding skipped variable pertaining to driving under the influence of that substance was assigned a code of 91 to indicate that the respondents were lifetime nonusers of that substance (Section 2.4.2). For example, if respondents never used alcohol, then the edited variable DRVINALCO (corresponding to question SP06A) was assigned a code of 91.
- If respondents last used the relevant substance more than 12 months prior to the interview, then the corresponding skipped variable was assigned a code of 93 to indicate that respondents did not use that substance in the past 12 months (Section 2.4.2).

If respondents were not asked the question about driving under the influence of alcohol, marijuana, hallucinogens, or inhalants, and the corresponding recency variable had been assigned a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), then the edited variable for driving under the influence retained a code of 98 (blank). There is a procedure in place to assign bad data codes (Section 2.3) to the variables for driving under the influence of alcohol, marijuana, hallucinogens, or inhalants if the corresponding recency variable had been assigned a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) but data existed in the variable for driving under the influence. Data for driving under the influence in the past 12 months would indicate that these respondents had used the corresponding

substance in that period. This edit could be implemented for the variable for driving under the influence of hallucinogens (DRVINHALL, corresponding to question SP06E) if there was a response of “don’t know” or “refused” in a consistency check between “parent” and “child” hallucinogen questions (Section 6.2.2.5). In this situation, the affected variables for parent and child hallucinogen recency would have been edited to a value of 9 but the CAI program continued to treat the respondents as past year users. This edit was needed for one case in 2023. In addition, if variables for a given drug had been assigned codes for bad data because of patterned responses (Section 2.3.2), then the corresponding variable for driving under the influence was assigned a code for bad data.

9.3.2.2 Driving under the Influence of Cocaine, Heroin, or Methamphetamine

Similar logic to that described in Section 9.3.2.1 was implemented for determining when respondents had legitimately skipped out of the corresponding questions for driving under the influence of cocaine, heroin, or methamphetamine because of nonuse in their lifetime or in the past 12 months. As noted previously, edits to the variables for driving under the influence of these substances also took into account data from the special drugs section of the interview. Thus, for example, if respondents reported lifetime nonuse of cocaine, heroin, or methamphetamine in that substance’s section of the interview, but the corresponding needle recency variable from the special drugs section had been edited to indicate use with a needle more than 12 months ago, then the skipped variable for driving under the influence (e.g., DRVINCOCN, corresponding to question SP06C) was coded as 93 (DID NOT USE IN THE PAST 12 MONTHS) instead of 91 (NEVER USED) as part of the editing.

As was also indicated previously, respondents who reported in sections of the interview for cocaine, heroin, or methamphetamine that they last used these substances more than 12 months ago but who reported use at some point in the past 12 months in the special drugs questions were asked the corresponding questions for driving under the influence. For example, if respondents reported in the cocaine section of the interview that they last used cocaine “more than 12 months ago,” but they reported using cocaine with a needle in the past 30 days or more than 30 days ago but within the past 12 months, then data from question SP06C were retained in the edited variable DRVINCOCN.

As for the edits described in Section 9.3.2.1, if respondents were not asked the question about driving under the influence of cocaine, heroin, or methamphetamine, and the corresponding recency variable had been assigned a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), then the edited variable for driving under the influence retained a code of 98 (blank). The edited variable for driving under the influence also retained a code of 98 (blank) if respondents reported that they last used cocaine, heroin, or methamphetamine more than 12 months ago in the corresponding recency variable but had missing data for recency questions for these substances in the special drugs section.

For driving under the influence of cocaine or methamphetamine in the past 12 months, the edited variables retained a code of 98 (blank) if respondents reported that they last used these drugs more than 12 months ago but had missing data from the special drugs section for when

they last used these drugs with a needle. Similarly, respondents who reported in the heroin recency (HERREC) that they last used heroin more than 12 months ago but had missing data from special drugs for smoking, sniffing, or using heroin with a needle were coded as 98 (blank) for driving under the influence of heroin.

9.3.2.3 Driving under the Influence of Alcohol Only

As noted previously, respondents were asked question SP06H about driving under the influence of only alcohol if they reported in question SP06A that they drove under the influence of alcohol in the past 12 months and they also reported that they drove under the influence of marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine in that same period. The edits described below were implemented for the variable DRVINALON for driving under the influence of only alcohol when respondents were determined to have skipped out of question SP06H.

- If respondents were lifetime nonusers of alcohol, then DRVINALON was assigned a code of 91 (NEVER USED ALCOHOL). In this situation, both DRVINALON and DRVINALCO (drove under the influence of alcohol in the past 12 months) were coded as 91.
- If respondents last used alcohol more than 12 months prior to the interview, then DRVINALON was assigned a code of 93 (DID NOT USE ALCOHOL IN THE PAST 12 MONTHS). In this situation, both DRVINALON and DRVINALCO were coded as 93.
- If respondents reported that they did not drive under the influence of alcohol in the past 12 months (DRVINALCO = 2), then DRVINALON was assigned a code of 99 (LEGITIMATE SKIP) (Section 2.4.2). This edit was implemented regardless of whether respondents reported that they drove under the influence of other substances (i.e., marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine) in the past 12 months.
- If respondents reported that they drove under the influence of alcohol in the past 12 months (DRVINALCO = 1), but they did not drive under the influence of other substances in that period (including situations where respondents did not use these substances in the past 12 months or in their lifetime), then DRVINALON was assigned a code of 99 (LEGITIMATE SKIP). In this situation, respondents logically drove only under the influence of alcohol in the past 12 months.
- If DRVINALCO had missing data but respondents did not drive under the influence of other substances in the past 12 months, then DRVINALON was assigned a code of 99 (LEGITIMATE SKIP). In this situation, respondents were not eligible to be asked question SP06H because they did not drive under the influence of other substances in the past 12 months, despite having unknown information for whether they drove under the influence of alcohol in that period.

DRVINALON retained a code of 98 (blank) in the situations described below when question SP06H had been skipped.

- Use of alcohol in the past 12 months was unknown and either of the following applied to driving under the influence of other substances in that period: (1) respondents drove under the influence of other substances, or (2) driving under the influence of one or

more other substances also was unknown. In these situations, both DRVINALON and DRVINALCO were coded as 98 (blank).

- Respondents used alcohol in the past 12 months, DRVINALCO had missing data (e.g., question SP06A was answered as “don’t know” or “refused”), and the same conditions that were described in the preceding bullet applied to driving under the influence of other substances in that period.
- Respondents drove under the influence of alcohol in the past 12 months (DRVINALCO = 1), they did not report driving under the influence of other substances in that period, but information was unknown for whether respondents drove under the influence of some of these other substances.

9.3.3 Editing of Variables for State Marijuana Laws

Minimal editing was done to the variables MXMJPNLT, MXMJSURE, and MEDMJUSE, corresponding to questions SP07, SP08, and SP09, respectively. These variables pertained to knowledge about state marijuana laws and penalties. If respondents indicated in MXMJPNLT that they did not know the maximum legal penalty in their states for possession of an ounce or less of marijuana for personal use, or if they refused to answer this question, the edited variable MXMJSURE (regarding respondents’ degree of certainty about their answer to question SP07) was assigned a legitimate skip code.

In addition, information from FIPE4 (the state where respondents were living) was used to edit MXMJPNLT, MXMJSURE, and MEDMJUSE; interviewers completed FIPE4 for in-person respondents, and web respondents answered FIPE4 themselves. If a response to the FIPE4 question regarding the state where the respondent’s sampled dwelling unit (DU) was located was inconsistent with the sample data for that respondent, the variables MXMJPNLT, MXMJSURE, and MEDMJUSE were assigned bad data codes (85).¹⁵³ This latter edit was done because the state that respondents were asked about for MXMJPNLT, MXMJSURE, and MEDMJUSE was governed by the state that was reported in FIPE4. Hence, if a response to FIPE4 contained state information that was inconsistent with the sample data, the answers that respondents provided in questions SP07 through SP09 were deemed to be questionable.

9.4 Social Environment

As noted in Section 9.1, the social environment section was administered only to adults. This section included questions about involvement in criminal or potentially criminal activities, attitudes about adults trying marijuana once or twice, and religious involvement.

Minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes for respondents who were aged 12 to 17 (Section 2.4.2). Adults were asked all questions within the social environment section. As noted in Section 2.3.2.2, responses were replaced with codes for bad data if it was determined that an adult respondent had a patterned response of keying only “1” or only “2” to every question in the section, and the patterned response (1) started in the social environment section and continued

¹⁵³ Creation of the edited variable STATELOC from FIPE4 is discussed in detail in Chapter 4.

into subsequent sections or (2) started in prior sections and continued in the social environment section. Because of the brevity of this section, no editing was done to responses of only “1” or only “2” if they occurred only in the social environment section.

9.5 Youth Experiences

As noted in Section 9.1, the youth experiences section was administered only to respondents aged 12 to 17. This section included questions about school enrollment and related issues (e.g., opinions about the importance of assigned schoolwork) in the past 12 months, including homeschooling; other social and family characteristics (e.g., substance use behaviors of other students or friends, personal attitudes about substance use, parental attitudes about substance use); people with whom the youth could confide about a serious problem; exposure to alcohol- and other drug-related prevention messages in school or outside school; and personal behaviors (e.g., involvement in criminal or potentially criminal activities, involvement in extracurricular activities) that might be associated positively or negatively with the use of alcohol or other drugs. The youth experiences section also included questions about youths’ religious involvement in the past 12 months and opinions about religious issues.

Minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes based on the CAI routing logic. That included (1) assignment of legitimate skip codes to variables in the entire section for respondents who were aged 18 or older and (2) assignment of legitimate skip codes to youths’ data based on routing logic within the youth experiences section.

As noted in Section 2.3.2.2, responses in the youth experiences section were replaced with bad data codes when respondents keyed all answers as “1” (or multiple-digit responses of “1,” where applicable) or all answers as “2” to every question that they were asked within a section. For respondents to key all possible items as “2” in the youth experiences section would mean that they keyed this response to 41 items in a row, including that they had engaged in all “problem” activities (such as getting into fights, carrying a handgun, or selling illegal drugs) exactly “1 or 2 times” in the past year. These respondents who keyed “2” wherever possible also would have participated in exactly two school-based activities, two community-based activities, two church- or faith-based activities, and two other kinds of activities in the past year.

Some special issues were encountered in editing the variables corresponding to question YE22, an “enter all that apply” question (Section 2.4.4) that asked about people to whom youths could turn to if they had a serious problem. Specifically, youths were asked to enter all the different types of people to whom they could turn to (e.g., a parent, a friend). This question also included a response category for youths who felt that there was no one they could talk to about a serious problem. If the entire list of responses was blank (e.g., if a youth broke off the interview before getting to these questions), the edited variables retained a code of 98 (blank).

Youths could indicate that there was no one they could talk to about a serious problem but then indicate that they could talk to one or more of the people or types of people in the list from question YE22. In this situation, the variable pertaining to the first item in the list (“There is nobody I can talk to about a serious problem”) was assigned a code of 11 (if that response was

chosen along with another response from the list). Similarly, a code of 11 was assigned to the edited relationship variables (e.g., my mom, my dad) when they were chosen along with the response that there was nobody that the youth could talk to.

9.5.1 Suicidal Thoughts and Behavior among Youths

Questions about suicidal thoughts and behavior among youths have been included in the NSDUH questionnaire since Quarter 4 of 2020. These questions mirrored the questions asked of adult respondents in the mental health questionnaire section. All youths were asked whether they had serious thoughts of suicide, made suicide plans, or attempted suicide in the past 12 months. Youths who reported suicidal thoughts or behavior in the past 12 months also were asked whether these thoughts or behavior were because of coronavirus disease 2019 (COVID-19). [Table 9.1](#) lists these questions regarding suicidal thoughts and behaviors among youths, the names of the corresponding edited variables, and descriptions for these questions.

Table 9.1 Mapping of Edited Variables for Suicidal Thoughts and Behavior among Youths

Question	Edited Variable	Description
YSUI01	YUSUITHK	Seriously thought about trying to kill yourself in the past 12 months.
YCOV9	YUCOSUITHK	Was this (serious thoughts of suicide) because of the COVID-19 pandemic?
YSUI02	YUSUIPLN	Made any plans to kill yourself in the past 12 months.
YCOV10	YUCOSUIPLN	Was this (suicide plans) because of the COVID-19 pandemic?
YSUI03	YUSUICTRY	Tried to kill yourself in the past 12 months.
YCOV11	YUCOSUICTRY	Was this (suicide attempts) because of the COVID-19 pandemic?
YSUI04	YUSUICMED	Got medical attention from a doctor or other health professional as a result of an attempt to kill yourself in the past 12 months.
YSUI05	YUSUIHOSP	Stayed in a hospital overnight or longer because you tried to kill yourself.

COVID-19 = coronavirus disease 2019.

An important aspect of processing these variables is assigning legitimate skip codes where relevant. As for other edited variables in the youth mental health service utilization section, legitimate skip codes were assigned to all adult respondents. Legitimate skip codes also were assigned to youth respondents who were skipped out of questions that did not apply. For example, respondents who did not report having serious thoughts of suicide in question YSUI01 were not asked follow-up question YCOV9 (if the serious thoughts of suicide were because of COVID-19). If respondents answered YSUI01 as “no,” the edited variable YUCOSUITHK (corresponding to YCOV9) was assigned a legitimate skip code.

9.6 Parenting Experiences

The parenting experiences section was intended to be administered only in DUs where (1) two people had been selected for an interview, (2) a 12- to 17-year-old had been selected for an interview (regardless of whether the youth completed the interview), and (3) the respondent being interviewed was the parent or legal guardian of the 12- to 17-year-old who also was

selected for an interview. Editing of the parenting experiences data involved editing the checkpoint variables (FIPE1, FIPE2, and FIPE3 for in-person respondents and FIPE1, YOUTH_SELECTED, and FIPE3B for web respondents).¹⁵⁴ Variables in the parenting experiences section were edited based on the final values assigned to the edited FIPE variables.

The checkpoint variables and questions for in-person and web respondents are described below. Section 9.6.1 explains how the information collected from in-person and web respondents is combined to create the edited checkpoint variables.

For in-person interviews, these checkpoint items were completed by interviewers toward the beginning of the interview. Questions FIPE1, FIPE2, and FIPE3 are shown below. Questions appeared in ALL CAPS to indicate to interviewers that questions were not to be read to respondents.

FIPE1 INTERVIEWER: WERE 2 PERSONS SELECTED FOR AN INTERVIEW AT THIS SDU?

- 1 YES
- 2 NO

FIPE2 [IF FIPE1 = 1 AND CURNTAGE = 18 OR OLDER] INTERVIEWER: WAS A 12 - 17 YEAR OLD CHILD SELECTED FOR AN INTERVIEW AT THIS SDU?

- 1 YES
- 2 NO

FIPE3 [IF FIPE2 = 1] INTERVIEWER: IS **THIS** RESPONDENT THE PARENT OR LEGAL GUARDIAN OF THE 12 - 17 YEAR OLD CHILD WHO WAS SELECTED FOR AN INTERVIEW? (VERIFY THIS WITH THE RESPONDENT IF YOU ARE UNSURE.)

- 1 YES
- 2 NO

For web respondents, FIPE1 and YOUTH_SELECTED were preloaded “behind the scenes” from sample selection information determined in the household screener.

- FIPE1 was coded as 1 (yes) if the household screener indicated that two people in the DU were selected for an interview or as 2 (no) otherwise.
- YOUTH_SELECTED was defined as 0, 1, or 2 for all web respondents according to the number of youths aged 12 to 17 who were selected for an interview.

FIPE3B (shown below) was asked at the beginning of the parenting experiences section for web respondents. RESAGE (for respondent age) in the specifications for FIPE3B was coded as 1 if the respondent’s age from the household screener was defined as being 18 years or older.

¹⁵⁴ “Checkpoint” refers to an item about the location of the sampled dwelling unit (DU) or the characteristics of the sample within the sampled DU. These checkpoints were used to determine what respondents were asked in subsequent sections of the interview.

Thus, FIPE3B was asked if the respondent was aged 18 or older and the other person selected at the DU was aged 12 to 17.

FIPE3B [IF RESAGE = 1 AND FIPE1 = 1 AND YOUTH_SELECTED = 1] Are you the parent or legal guardian of the 12-17 year old in this household who was also selected for an interview?

- 1 Yes
- 2 No
- DK/REF

The audio computer-assisted self-interviewing (ACASI) “lockout” feature for in-person interviews that was described in Section 2.4.6 did not affect the editing of checkpoint data but did affect the editing of parenting experiences data. Issues associated with this ACASI “lockout” that are relevant to the editing of the parenting experiences variables are described in Section 9.6.2.

9.6.1 Editing of the Checkpoint Variables

The edited variables SEL2PER, SEL1217, and SELPRNT were edited for consistency with the pair-selection and pair-respondent sample variables (PAIRSEL and PAIRRESP, respectively). Final values in the edited checkpoint variables were critical for determining whether respondents were eligible to be asked the parenting experiences questions. [Table 9.2](#) shows how these variables were created to combine the information collected from in-person and web interviews.

Table 9.2 Parenting Experiences Checkpoint Questions by Interview Mode

Edited Variable	Questionnaire Variable	In-Person Survey	Web Survey
SEL2PER	FIPE1 (Were two people selected?)	Completed by interviewer	Preloaded from household screener data
SEL1217	FIPE2 (Was a 12- to 17-year-old selected?)	Completed by interviewer	N/A
SEL1217	YOUTH_SELECTED (How many 12- to 17-year-olds were selected?)	N/A	Preloaded from household screener data. To correspond to in-person data from FIPE2 for further processing, the variable was recoded for adults as 1 (yes) if 1 or 2 youths were selected and as 2 (no) if 0 youths were selected.
SELPRNT	FIPE3 (Is respondent the parent for the selected youth?)	Completed by interviewer ¹	N/A
SELPRNT	FIPE3B (Is respondent the parent for the selected youth?)	N/A	Answered by respondents ¹

N/A = not applicable.

¹ Adult respondents were asked the parenting experiences questions if FIPE3 or FIPE3B was answered as “yes.”

9.6.1.1 Editing of SEL2PER

The FIPE1 variable, which was comparable for in-person and web respondents, was edited for consistency with PAIRSEL. This checkpoint pertained to whether two people were selected for an interview at a DU. There were no situations in 2023 when two people were interviewed at a DU without two people having first been selected. Therefore, editing FIPE1 for in-person and web respondents involved reviewing only information on the number of people selected for an interview at that DU based on PAIRSEL.

If the pair-selection data indicated that two people were selected from that DU, then FIPE1 should have been answered as “yes” in in-person interviews or preset to “yes” in web interviews. Therefore, if the pair-selection data indicated that two people were selected and FIPE1 was not answered or precoded as “yes,” then a code of 3 (Yes LOGICALLY ASSIGNED) was assigned to the edited FIPE1 variable (SEL2PER). Similarly, if the pair-selection data indicated that only one person was selected from that DU, then FIPE1 should have been answered or precoded as “no.” Therefore, if the pair-selection data indicated that only one person was selected and FIPE1 was not answered or precoded as “no,” then a code of 4 (No LOGICALLY ASSIGNED) was assigned to SEL2PER. Assignment of legitimate skip codes for editing of SEL1217 and SELPRNT when SEL2PER indicated that a 12- to 17-year-old was not selected for an interview is discussed below.

9.6.1.2 Editing of SEL1217

To facilitate processing of SEL1217, the YOUTH_SELECTED variable from web respondents was recoded as a yes/no variable to be consistent with the FIPE2 variable from in-person data.

- The recoded version of YOUTH_SELECTED was treated as not applicable (i.e., blank) if FIPE1 indicated that only one person was selected or if the respondent was aged 12 to 17.
- The recoded version of YOUTH_SELECTED was coded as 1 (yes) if the respondent was aged 18 or older and a second person aged 12 to 17 was selected.
- The recoded version of YOUTH_SELECTED was coded as 2 (no) if the respondent was aged 18 or older and no youths aged 12 to 17 were selected.

After data for web interviews were recoded, SEL1217 was assigned legitimate skip codes if SEL2PER indicated that only one person was selected. SEL1217 was coded as 99 (LEGITIMATE SKIP) if FIPE2 for in-person respondents was blank or if the recoded version of YOUTH_SELECTED for web respondents was treated as not applicable. Otherwise, data were overwritten with codes of 89 (LEGITIMATE SKIP Logically assigned) if SEL2PER = 4 (No LOGICALLY ASSIGNED) and data existed (i.e., yes or no).

If two people were selected for an interview, SEL1217 was edited for consistency with PAIRSEL, PAIRRESP, and the age of the respondent. This checkpoint pertained to whether a 12- to 17-year-old was selected for an interview at the DU, regardless of whether the selected youth responded. Edits of SEL1217 involved review of both the pair-selection data (PAIRSEL) and the

pair-respondent data (PAIRRESP) in case either indicated that a 12- to 17-year-old was selected or interviewed.

For in-person data, the age of the respondent was taken into account because interviewers were skipped past the FIPE2 checkpoint if respondents were aged 12 to 17. As noted previously, this checkpoint was treated as not applicable for web interviews if the respondent was aged 12 to 17. Therefore, SEL1217 was assigned legitimate skip codes when the respondent was a youth, regardless of interview mode.

The remaining edits for SEL1217 were implemented when the respondent was an adult. If both PAIRSEL and PAIRRESP indicated that a 12- to 17-year-old was neither selected nor interviewed, it reasonably could be inferred that SEL1217 should indicate that a 12- to 17-year-old was not selected. If SEL1217 did not indicate that a 12- to 17-year-old was not selected based on initial edits, the edit assigned a code to SEL1217 to indicate that a response of "no" was logically inferred. Implementation of this edit included situations in which the pair-selection data indicated that a 12- to 17-year-old was not selected and that a completed interview was obtained from only one respondent who was not aged 12 to 17, regardless of whether PAIRSEL and PAIRRESP were totally consistent. For example, if the pair-selection data indicated that an 18- to 25-year-old and a 26- to 34-year-old were selected, but a single interview was obtained from a 35- to 49-year-old, then the pair-selection and pair-respondent data were not totally consistent, but neither would suggest that a 12- to 17-year-old should have been selected.

If either PAIRSEL or PAIRRESP indicated that a 12- to 17-year-old was selected or interviewed, it was inferred that the data should have indicated selection of someone in that age group. Therefore, if selection of a 12- to 17-year-old had not already been reported, a special code was assigned to SEL1217 to indicate the logical inference that someone in that age group had been selected. Implementation of this edit included the following situations: (1) PAIRSEL indicated that a 12- to 17-year-old was selected, and PAIRRESP indicated that an interview was obtained from a 12- to 17-year-old, regardless of whether PAIRSEL and PAIRRESP matched exactly (e.g., a 12- to 17-year-old and a 26- to 34-year-old were selected, but interviews were obtained from a 12- to 17-year-old and a 35- to 49-year-old); and (2) PAIRSEL indicated that a 12- to 17-year-old was selected, but a single interview from an adult was obtained at the DU, regardless of whether the adult category from PAIRSEL matched the adult category in PAIRRESP (e.g., a 12- to 17-year-old and a 26- to 34-year-old were selected, but a single interview was obtained from a 35- to 49-year-old). In the latter situation, the respondent result (from PAIRRESP) was not totally consistent with the expected result based on the pair selection, but the information from PAIRRESP did not directly contradict the indication from PAIRSEL that a 12- to 17-year-old was selected.

Special codes were assigned to SEL1217 if PAIRSEL and PAIRRESP disagreed (i.e., only one of these variables indicated the selection or interview of a 12- to 17-year-old within a pair). A code of 11 was assigned to SEL1217 when the original data indicated that a 12- to 17-year-old had been selected, and a code of 12 was assigned when the original data indicated that someone in that age group had not been selected.

Suppose, for example, that PAIRSEL indicated that a 12- to 17-year-old and a 35- to 49-year-old were selected for the interview, but PAIRRESP indicated that an 18- to 25-year-old and a 35- to 49-year-old were actually interviewed, and selection data indicated that a 12- to 17-year-old was selected for an interview. In this situation, the data were consistent with the ages of those who were selected but were not consistent with the ages provided by the respondents. Therefore, the edited variable SEL1217 would be set to a value of 11 in this example.

This edit preserved the original information on whether a 12- to 17-year-old was selected, but it also denoted that an inconsistency existed between PAIRSEL and PAIRRESP. This edit was designed to also preserve any parenting experiences data when checkpoint variables indicated that a 12- to 17-year-old was selected and the adult respondent was the parent or guardian of the selected youth but there was an inconsistency between PAIRSEL and PAIRRESP. When this inconsistency occurred, analysts had discretion about whether to use parenting experiences data in an analysis.

9.6.1.3 Editing of SELPRNT

This variable pertained to whether the respondent at the DU was the parent or legal guardian of the 12- to 17-year-old who also was selected to be interviewed at that DU. As noted previously, FIPE3 was answered by interviewers at the beginning of in-person interviews. FIPE3B was answered by adult respondents at the beginning of the parenting experiences section in web interviews. FIPE3 and FIPE3B did not apply to respondents who were aged 12 to 17. Therefore, when these questions had been skipped because respondents were aged 12 to 17, the edited variable SELPRNT was assigned a legitimate skip code. When the edited SEL1217 variable indicated that a 12- to 17-year-old was not selected, including situations in which the edits inferred that a 12- to 17-year-old was not selected, legitimate skip codes also were assigned to the edited variable SELPRNT. If FIPE3 for in-person respondents or FIPE3B for web respondents was blank, then SELPRNT was assigned a code of 99 (LEGITIMATE SKIP). Otherwise, data in FIPE3 or FIPE3B were overwritten with a code of 89 (LEGITIMATE SKIP Logically assigned).

No further editing of SELPRNT was done when PAIRSEL indicated that two people, including a 12- to 17-year old, were selected and PAIRRESP had some result *other than* that of two adults having been interviewed at that DU. The rationale for this approach was that this variable was based on the age of the actual respondent, provided that a 12- to 17-year-old was selected. For example, if PAIRSEL indicated that a 12- to 17-year-old and a 26- to 34-year-old were selected, but a 12- to 17-year-old and a 35- to 49-year-old were interviewed, and data indicated that the adult respondent was the parent of the youth who was selected, then that 35- to 49-year-old respondent indeed may have been a parent or legal guardian of the youth who was selected. This principle also would apply if the selected youth did not respond. Therefore, any data that were present in the parenting experiences section would be preserved.

In contrast, the following situations could have occurred when original data corresponding to SELPRNT was inconsistent with either PAIRSEL or PAIRRESP: (1) PAIRSEL indicated that a youth/adult pair was selected, but two adult interviews were obtained at that DU, or (2) PAIRRESP indicated that a youth/adult pair was interviewed, but PAIRSEL indicated that an

adult/adult pair was selected. When either inconsistency occurred, a code of 11 was assigned to SELPRNT when the questions were originally answered as “yes,” and a code of 12 was assigned when the questions were originally answered as “no.”

Suppose, for example, that the following pattern occurred in the data:

- PAIRSEL indicated that an 18- to 25-year-old and a 35- to 49-year-old were selected for the interview.
- PAIRRESP indicated that a 12- to 17-year-old and a 35- to 49-year-old were actually interviewed (rather than an 18- to 25-year-old and a 35- to 49-year-old).
- Data from FIPE2 or YOUTH_SELECTED indicated that a 12- to 17-year-old was selected.
- FIPE3 or FIPE3B was answered as “yes,” indicating that the adult respondent who was aged 35 to 49 was the parent of the selected youth.

In this situation, the answer from FIPE3 or FIPE3B was consistent with PAIRRESP but not with PAIRSEL. Furthermore, based on who was interviewed at the DU, the 35- to 49-year-old may be the parent of the 12- to 17-year-old who also was interviewed at that DU. In this situation, the edited SELPRNT variable would be set to a value of 11 to denote that this type of inconsistency occurred. Again, this edit preserved all parenting experiences data, especially in situations in which an adult/child respondent pair was obtained.

9.6.2 Editing of Variables in the Parenting Experiences Section

Variables in the parenting experiences section were edited according to the final values assigned to SEL2PER, SEL1217, and SELPRNT based on the edits described in Section 9.6.1.¹⁵⁵ For in-person respondents, if these three variables indicated that the respondent was not eligible to be administered the parenting experiences questions, then the edit assigned the appropriate legitimate skip codes to the parenting experiences variables. For web respondents, the edit assigned legitimate skip codes to the remaining parenting experiences variables following SELPRNT.

Implementation of this edit included replacing values of “blank” with legitimate skip codes when a code of 12 had been assigned to SELPRNT and the parenting experiences section (or the remainder of the section following SELPRNT for web respondents) had been skipped. The rationale for this latter edit was that even if FIPE3 or FIPE3B was answered as “no” when PAIRSEL and PAIRRESP were inconsistent, then the adult respondent at the DU still may not have been the parent or legal guardian of the youth who also was selected for an interview at that DU.

Conversely, if a respondent had been skipped out of the parenting experiences section, and the edited FIPE variables SEL2PER, SEL1217, and SELPRNT indicated that the respondent potentially was eligible to be administered the parenting experiences questions (i.e., the respondent skipped the section based on the original answers to the FIPE questions, but other

¹⁵⁵ For web interviews, FIPE3B (corresponding to SELPRNT) was the first question in the parenting experiences section.

data suggested that the respondent may have been eligible to be asked these questions), then the edited parenting experiences variables retained a value of "blank." For example, if FIPE2 or YOUTH_SELECTED indicated that a 12- to 17-year-old had not been selected, but it was inferred for SEL1217 that a 12- to 17-year-old was selected (i.e., SEL1217 = 3), then FIPE3 or FIPE3B and the parenting experiences questions (or remaining questions) also would have been skipped. In this situation, the respondent's eligibility to be administered the parenting experiences questions could not be determined because the final checkpoint question was not asked. Therefore, it could not be determined whether the respondent should have been asked the parenting experiences questions or the questions should have been skipped.

As noted previously, the in-person interview included a "lockout" feature that did not allow interviewers or respondents to go back into the ACASI sections and change their answers once that section of the interview had been completed. For example, the parenting experiences section was skipped if interviewers had originally entered answers in FIPE1 through FIPE3 to indicate that respondents were not eligible to receive the parenting experiences questions (e.g., if FIPE1 indicated that two people were not selected at that DU). However, interviewers could go back to the beginning of the interview after respondents had been administered the ACASI sections and change FIPE1 through FIPE3 in a manner that indicated that respondents should have been asked the parenting experiences questions (i.e., two people were selected at the DU, a 12- to 17-year-old was selected, and the respondent was the parent of the selected 12- to 17-year-old). When this situation occurred, in-person respondents could no longer be routed back to the parenting experiences section. Consequently, the skipped parenting experiences variables had a final value of "blank."

This issue did not apply to web respondents because all web interview questions were self-administered. Hence, there was no need to "lock" web respondents out of previous sections. Regardless of interview mode, respondents were given instructions for how they could back up in the interview one question at a time. However, respondents were not given instructions for how they could jump from a later section back to the beginning of the interview.

The parenting experiences variables also could have a final value of "blank" if potentially eligible adult respondents provided usable information on their substance use (Section 2.2.3), but they stopped the interview before reaching the parenting experiences section (i.e., "break-offs"). Item nonresponse among adults is typically driven by survey break-offs.

Following the checkpoint questions (FIPE1, FIPE2, and FIPE3 for in-person respondents; FIPE1, YOUTH_SELECTED, and FIPE3B for web respondents), the parenting experiences questions were the same for in-person and web respondents. Therefore, remaining edits described in this section for parenting experiences variables were the same regardless of interview mode.

One set of variables within the parenting experiences section involved skip logic. Specifically, respondents were skipped out of question PE04 (length of most serious discussion about the dangers of tobacco/alcohol/other drug use) when question PE03 had a value of 1 (i.e., talked with child 0 times in the past year about the dangers of tobacco/alcohol/other drug use) or when PE03 was answered as "don't know" or "refused." Standard procedures for assigning legitimate skip codes (Section 2.4.2) or propagating refusal codes (Section 2.4.3) were

implemented in the edited version of question PE04 (PXSERDIS) depending on the response to PE03 (edited variable PXKIDYR).

Parents were asked to report the birth date of the youth who was selected for an interview at the DU (question PE01). However, the birth year that respondents could enter for the youth in question PE01 was restricted to ages that would be more consistent with selection of a 12- to 17-year-old (but also allowed for birth dates that would include 18-year-olds, in case a 17-year-old respondent recently had a birthday). Thus, respondents were prevented from entering birth dates that would be extremely inconsistent with selection of a 12- to 17-year-old (e.g., entry of the current interview year for the birth year).

The parenting experiences edit logic took into account the situation in which all remaining parenting experiences questions had been skipped because the respondent did not provide a date of birth for the selected youth in question PE01 or did not provide an age for the youth in question PE01B. The parenting experiences variables retained codes of "blank" when this pattern occurred. This edit makes the frequencies of legitimate skip values in PXCHCIG (corresponding to question PE02) and subsequent parenting experiences variables agree with the total count of codes of 2 or 12 (i.e., "no") or 89 or 99 (i.e., legitimate skip) in SELPRNT.

The CAI program also calculated an age for the youth who was selected for an interview based on the youth's date of birth (as reported by the parent) and the interview date at the start of the parenting experiences section. Respondents were asked to confirm this age (question PE01A). If parents did not confirm the age that the CAI program calculated for the youth, they were asked to provide a corrected age for the youth who was selected for an interview (question PE01B). Similarly, if respondents did not know or refused to report the date of birth of the selected youth, they were asked to report an age in question PE01B without having to indicate the youth's date of birth.

This information was captured in the edited variable PXCHLDAG. Specifically, PXCHLDAG contained the age based on the reported date of birth for the youth and the interview date (if respondents confirmed that this age was correct). Otherwise, PXCHLDAG contained the age supplied by the respondent from question PE01B. If respondents supplied a corrected age for the youth in question PE01B that was between 12 and 18 and it mismatched the age of the youth that was calculated from the birth date and interview date information, then the edited variables containing the birth date information for the youth (PXBMONTH, PXBDAY, and PXBYR) were assigned bad data values. If respondents answered question PE01B as "don't know" or "refused" when they were asked to provide a corrected age for the selected youth, then that response of "don't know" or "refused" was assigned to PXCHLDAG. In addition, if respondents answered question PE01A as "don't know" or "refused" when they were asked to confirm the age of the youth who was selected for the interview, then the age for the youth that the CAI program had previously calculated was retained in PXCHLDAG (see above). When values in parenting experiences variables had been set to bad data because SELPRNT was blank, PXCHLDAG also was assigned a code for bad data.

If a 12- to 17-year-old was selected for the second interview and the interview was completed, then a recoded variable, PXCMPAGE, was calculated as the absolute value of the difference

between PXCHLDAG (as reported by the parent) and the actual age of the second respondent (as reported by the second respondent during the interview). The difference in value is shown within defined categories (i.e., difference of 0 years in ages; difference of 1 year in ages; difference of 2 years in ages; difference of 2 to 3 years in ages; and difference of 5 years or more in ages). If the adult respondent answered “don’t know” or “refused” to the question about the youth’s date of birth, or if the youth’s date of birth information was set to bad data because of invalid dates, then these codes were reflected in PXCMPAGE.

For the majority of situations in which an interview was obtained from a 12- to 17-year-old, PXCMPAGE indicated no difference between the age based on the date of birth reported by the parent and the youth’s age that was recorded in the second interview at that DU. Nevertheless, analysts could use information about more extreme differences in ages as recorded by PXCMPAGE (e.g., a difference of 2 or more years between the two ages) to decide whether to use the parenting experiences data in an analysis. When the second interview was from an 18-year-old, PXCMPAGE was assigned a value of 18. When the second interview was from an adult older than age 18 (and the parent was supposed to be reporting about a 12- to 17-year-old), the edit program assigned a code of 50 to PXCMPAGE. Again, these codes were designed to give analysts discretion in using or disregarding parenting experiences data when the second interview at a DU came from an adult.

If a 12- to 17-year-old was supposed to be selected at a DU, but only the adult at that DU was interviewed, then PXCMPAGE was assigned a code of 93. This code was assigned because there were no data to corroborate the youth’s date of birth reported by the parent.

If the edited FIPE variables indicated that the respondent was not eligible to be administered the parenting experiences questions, then PXCMPAGE was assigned a code of 99 (LEGITIMATE SKIP). For example, PXCMPAGE was assigned a code of 99 (LEGITIMATE SKIP) if SELPRNT had been skipped because PAIRSEL indicated that only one person had been selected (SEL2PER = 2). PXCMPAGE also was assigned a code of 99 (LEGITIMATE SKIP) when the edited variable created from FIPE3 and FIPE3B was assigned a code of 12 because of an inconsistency between PAIRSEL and PAIRRESP and the parenting experiences section had been skipped (see above). Otherwise, if the parenting experiences section was all blank or PXCMPAGE was undefined for some other reason, then PXCMPAGE was assigned a code of 98. This code of 98 in PXCMPAGE meant “other missing” and was applied when SELPRNT was blank and parenting experiences data had been replaced with codes for bad data.

10. Editing and Imputation of Sections for Physical and Mental Health

10.1 Introduction

This chapter discusses the editing and imputation procedures for physical and mental health. Editing occurred in the following sections:

- health,
- mental health services utilization,
- adult mental health,
- adult or adolescent depression, and
- coronavirus disease 2019 (COVID-19).

Missing data for some variables related to adult mental health and adult depression were replaced using statistical imputation beginning in 2021. Missing data for some variables related to mental health services utilization were replaced using statistical imputation beginning in 2022. The imputation procedures for these variables are discussed in Section 10.7.

These questionnaire sections, which were self-administered for in-person respondents,¹⁵⁶ are included in this chapter according to their related topical content rather than in the sequential order in which they appeared to respondents during the interview. The order in which these sections were presented to respondents relative to other sections of the interview can be found in the computer-assisted interviewing (CAI) specifications for the 2023 National Survey on Drug Use and Health (NSDUH) that are available on the [NSDUH Data Collection](#) web page.

Beginning in 2022, the adult mental health service utilization and youth mental health service utilization sections were combined, updated, and renamed the mental health services utilization section. The revised section included terminology updates and additional treatment options. It was restructured to parallel the changes in the revised alcohol and drug use treatment section.

10.2 Health

The health section in the 2023 NSDUH covered a wide variety of health-related topics:

- pregnancy status for females aged 12 to 44,
- height and weight,
- utilization of medical services in the past 12 months,
- screening for substance use or talks with a healthcare professional about substance use in the context of healthcare visits in the past 12 months, and
- specific health conditions.

¹⁵⁶ Recall that all questions were self-administered for web respondents in 2023.

10.2.1 Pregnancy Status

Female respondents aged 12 to 44 were asked whether they were currently pregnant and, if so, the number of months that they had been pregnant. An important aspect of processing the pregnancy variables for females aged 12 to 44 involved assignment of legitimate skip codes, where relevant (Section 2.4.2). For example, males of any age and women over the age of 44 were assigned legitimate skip codes to the pregnancy variables. Similarly, if females aged 12 to 44 reported that they were not currently pregnant (PREGNANT = 2), legitimate skip codes were assigned to the variable pertaining to the number of months that they were pregnant (PREGMOS).

In the pregnancy variables, if women reported currently being pregnant, the allowable range for the number of months that they were pregnant ranged from 1 to 9 months. Thus, women who reported that they were currently pregnant were not allowed to report that they had been pregnant for “0” months.

10.2.2 Height and Weight

The 2023 NSDUH included questions for respondents to report their height and weight. Respondents were given four options for reporting their height: (1) in feet and inches, (2) in meters and centimeters, (3) in inches only, or (4) in centimeters only. The edited variable HTANSWER was a two-level variable that indicated whether respondents preferred to report their height in (1) English units (feet and inches or inches only) or (2) metric units (meters and centimeters or centimeters only).

Respondents could choose to report their weight in pounds or in kilograms. If female respondents aged 12 to 44 reported that they were currently pregnant (Section 10.2.1), they were asked to report their weight before they became pregnant. The edited variable WTANSWER was a two-level variable that indicated whether respondents preferred to report their height in (1) English units (pounds) or (2) metric units (kilograms).

A single variable for height in inches (HTINCHES) was created from respondents’ reports of their height. If respondents chose to report their height in feet and inches, their reports were converted to inches (e.g., 5 feet, 10 inches = 70 inches). If respondents chose to report their height in meters and centimeters, their answers first were converted to centimeters (e.g., 1 meter, 50 centimeters = 150 centimeters). Reported heights in centimeters (including heights in meters and centimeters that were converted to centimeters) were converted to heights in inches using the conversion factor of 1 centimeter = 0.393700787 inches.

The allowable ranges for reporting heights in inches (questions HLTH06A or HLTH06B) reflect whether respondents preferred to report their height in feet and inches or only in inches. Similarly, the allowable ranges for reporting heights in centimeters (questions HLTH08A or HLTH08B) reflect whether respondents preferred to report their height in meters and centimeters or only in centimeters. Respondents who chose to report their height in feet and inches could choose a number of feet ranging from 2 to 8 and a number of inches ranging from 0 to 11. Respondents who chose to report their height only in inches could choose a height

ranging from 24 to 96 inches. For metric units, respondents who chose to report their height in meters and centimeters could choose a number of meters ranging from 0 to 2 and a number of centimeters ranging from 0 to 99. Respondents who chose to report their height only in centimeters could choose a height ranging from 60 to 250 centimeters.

The allowable range for reporting weight in pounds (questions HLTH10 and HLTH13) was 40 to 999 pounds. Similarly, the allowable range for reporting weight in kilograms (questions HLTH12 and HLTH14) was 18.00 to 999.00 kilograms. As a consequence, codes for missing data for respondents' final weights were four digits in length.

HTINCHES could have missing values based on the types of responses that are described below.

- If respondents had missing data for their preferred way of reporting their height (edited variable HTANSWER), remaining questions for height were skipped. The corresponding code for missing data was assigned to HTINCHES (Section 2.2.1). For example, if HTANSWER = 94 (don't know), then HTINCHES was assigned a code of 994 because heights of 90 or more inches were allowed in the data.
- If respondents preferred to report their height in feet and inches or in meters and centimeters, it was possible for them to report the major unit for their height (e.g., feet) but to have a missing value for the minor unit (e.g., inches). The corresponding code for missing data was assigned to HTINCHES. For example, if respondents reported being 5 feet tall but did not know the number of additional inches beyond 5 feet for their height, then HTINCHES was assigned a code of 994.

A single variable for weight in pounds (WTPOUNDS) was created from respondents' reports of their weight (or the weight of pregnant females before they became pregnant). Reported weights in kilograms were converted to weights in pounds using the conversion factor of 1 kilogram = 2.20462262 pounds.

As for HTINCHES, relevant codes for missing data were assigned to WTPOUNDS according to the type of nonresponse. For example, if WTANSWER had a code of 94 (don't know), then WTPOUNDS was assigned a code of 9994. As noted previously, codes for missing data in WTPOUNDS were four digits in length to account for the allowable ranges for weight.

10.2.3 Utilization of Medical Services

The 2023 NSDUH questionnaire included questions about respondents' utilization of hospital emergency room services, overnight inpatient hospitalizations, virtual or telehealth services (over the phone or through video), and respondents' outpatient medical visits in the past 12 months. An important aspect of processing these healthcare utilization variables involved assignment of legitimate skip codes where relevant (Section 2.4.2). Specifically, respondents who did not report that they were hospitalized overnight in the past 12 months (edited variable INHOSPYR) were not asked for the number of times they were hospitalized in that period (edited variable NMNGTHSP). If respondents reported that they were not hospitalized overnight in the past 12 months (INHOSPYR = 2), the variable NMNGTHSP was assigned a legitimate skip

code. If respondents refused to report whether they were hospitalized overnight in the past 12 months (INHOSPYR = 97), that refusal was propagated onto NMNGTHSP (Section 2.4.3).

In addition, question HLTH19 (edited variable NMVSOPT), regarding the number of times that respondents visited healthcare professionals in outpatient settings, was a continuous variable. If respondents did not know or refused to report the number of times they visited healthcare professionals in outpatient settings in the past 12 months, they were asked in a follow-up question (HLTH19DK) to give their “best guess” of the number of outpatient visits that they had in the past 12 months. To encourage respondents to report a number of visits, HLTH19DK was categorical and included ranges for respondents to report their number of visits (e.g., 2-3 times, 4-5 times). HLTH19DK also had response categories of 0 (“I have not visited a healthcare professional in the past 12 months”) and 1 (i.e., 1 time). If NMVSOPT indicated a specific number of outpatient visits in the past 12 months, then the edited variable NMVSOEST corresponding to HLTH19DK was assigned a legitimate skip code.

For the questions about the number of nights that respondents were inpatients in a hospital or the number of outpatient visits in the past 12 months, allowable values that respondents could report ranged from 1 to 366. No editing was done to the variables NMNGTHSP or NMVSOPT when respondents reported that they had spent 365 nights in a hospital in the past 12 months, or that they visited a healthcare professional in an outpatient setting 365 or 366 times, respectively. In addition, no editing was done if the sum of the number of nights as an inpatient and the number of outpatient visits in the past 12 months was greater than 365. However, if respondents reported spending 366 nights as an inpatient in a hospital in the past 12 months, NMNGTHSP was set to 365. Because NMVSOPT refers to the number of outpatient *visits* rather than the number of days, no editing was done to NMVSOPT if respondents reported 366 outpatient visits in the past 12 months.

10.2.4 Screening for Substance Use or Talks about Substance Use with a Healthcare Professional

The health section in 2023 included questions HLTH20A through HLTH20C regarding whether respondents who had healthcare visits (including virtual or telehealth services) in the past 12 months were asked by a doctor or other healthcare professional about their tobacco, alcohol, or other substance use in the context of these medical visits (i.e., screening for substance use), regardless of whether they reported use of tobacco, alcohol, or illicit drugs. Respondents also were asked questions HLTH21 through HLTH23 about talks that they had with a doctor or other healthcare professional about their use of tobacco, alcohol, marijuana, or other illicit drugs if they reported healthcare visits (including virtual or telehealth services) in the past 12 months and they reported use of tobacco products, use of alcohol, or use of specific illicit drugs (i.e., marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, or methamphetamine), respectively.

Respondents were classified as having had healthcare visits in the past 12 months for the routing to questions HLTH20A through HLTH23 if any of the following occurred:

- Respondents reported in question HLTH16 (edited variable NMERTMT) that they had been treated in a hospital emergency room one or more times in the past 12 months.
- Respondents answered question HLTH17 (edited variable INHOSPYR) as “yes,” indicating that they stayed overnight or longer as an inpatient in a hospital in the past 12 months.
- Respondents reported in question HLTH19 (edited variable NMVSOPT) or in question HLTH19DK (edited variable NMVSOEST) that they had one or more visits to a healthcare professional in an outpatient setting in the past 12 months.
- Respondents answered question HLTHTELE (edited variable TELEAPTYR) as “yes,” indicating that they talked to a doctor or other healthcare professional about their own health over the phone or through video in the past 12 months instead of going to an in-person appointment.

Respondents were classified as not having had any healthcare visits in the past 12 months when the following occurred:

- Respondents reported zero emergency room visits in HLTH16.
- Respondents answered both HLTH17 and HLTHTELE as “no.”
- Respondents reported zero healthcare visits in HLTH19 or HLTH19DK.

Thus, an important aspect of processing the variables pertaining to screening for substance use and talks about substance use with healthcare professionals in healthcare settings involved assignment of legitimate skip codes where relevant (Section 2.4.2). Further details are provided elsewhere in this section for assignment of legitimate skip codes, depending on the questions.

10.2.4.1 Screening for Substance Use

As discussed previously, eligibility for being asked questions HLTH20A, HLTH20B, and HLTH20C was based only on reports of in-person or virtual healthcare visits in the past 12 months, regardless of whether respondents used tobacco, alcohol, or other substances. HPUSETOB, HPUSEALC, and HPUSEDRG were assigned a code of 99 (LEGITIMATE SKIP) if respondents reported no healthcare visits in the past 12 months, as defined in the previous section. HPUSETOB, HPUSEALC, and HPUSEDRG retained missing values if the questions determining healthcare visits had been skipped because respondents did not report having any in-person or virtual healthcare visits in the past 12 months but (1) NMERTMT, INHOSPYR, or TELEAPTYR had missing values (i.e., “don’t know” or “refused”) or (2) both NMVSOPT and NMVSOEST had missing values.

10.2.4.2 Talks about Tobacco Use

Respondents who had healthcare visits in the past 12 months and reported past year use of cigarettes, smokeless tobacco, or cigars, or past month use of pipe tobacco¹⁵⁷ were eligible to be asked question HLTH21 (HPQTTTOB), regarding whether a doctor or other healthcare professional advised these respondents to quit smoking cigarettes or to stop using other tobacco products. Data in HPQTTTOB were edited according to this skip logic as described below.

- If question HLTH21 had been skipped and respondents never used cigarettes, smokeless tobacco, cigars, or pipe tobacco, then HPQTTTOB was assigned a code of 91, where 91 = NEVER USED TOBACCO PRODUCTS. This code was assigned regardless of whether respondents reported any healthcare visits in the past 12 months.
- If question HLTH21 had been skipped because respondents did not report healthcare visits in the past 12 months and the most recent use of any tobacco products was more than 12 months ago (which could include reports that respondents never used some types of tobacco products),¹⁵⁸ then HPQTTTOB was assigned a code of 93, where 93 = DID NOT USE TOBACCO IN THE PAST 12 MONTHS.
- If question HLTH21 had been skipped because respondents had no healthcare visits in the past 12 months but they used any tobacco products in the past 12 months,¹⁵⁹ then HPQTTTOB was assigned a code of 99, where 99 = LEGITIMATE SKIP.
- Precedence was given to assigning codes of 91 or 93 to HPQTTTOB ahead of assigning a code of 99. Thus, respondents who did not use tobacco either in their lifetime or in the past 12 months and reported no healthcare visits were assigned codes for no tobacco use in the appropriate time period.

HPQTTTOB retained missing values if HLTH21 had been skipped because of the following:

- respondents reported healthcare visits in the past 12 months but it was unknown whether respondents were users of tobacco products in the past year; or
- respondents reported past year use of one or more tobacco products but it was unknown whether respondents had any healthcare visits in the past 12 months.

10.2.4.3 Talks about Alcohol Use

For question HLTH22, respondents who had healthcare visits in the past 12 months and reported past year use of alcohol were eligible to be asked the question. Because HLTH22 was an “enter all that apply” question, separate variables were created that corresponded to the response options in HLTH22. When HLTH22 was skipped, variables corresponding to HLTH22 were edited as described below.

¹⁵⁷ There is no question in NSDUH to identify respondents who used pipe tobacco more than 30 days ago but within the past 12 months.

¹⁵⁸ For brevity, use more than 12 months ago includes reports that respondents did not use pipe tobacco in the past 30 days and there were no reports of past year use of other tobacco products.

¹⁵⁹ Use in the past 12 months includes reports that respondents used pipe tobacco in the past 30 days.

- If question HLTH22 had been skipped and respondents never used alcohol, then the variables corresponding to HLTH22 were assigned a code of 91, where 91 = NEVER USED ALCOHOL. This code was assigned regardless of whether respondents reported any healthcare visits in the past 12 months.
- If question HLTH22 had been skipped because respondents last used alcohol more than 12 months prior to the interview date, then the edited variables corresponding to HLTH22 were assigned a code of 93, where 93 = DID NOT USE ALCOHOL IN THE PAST 12 MONTHS. Again, this code was assigned regardless of whether respondents reported any healthcare visits in the past 12 months.
- If question HLTH22 had been skipped because respondents had no healthcare visits in the past 12 months but they used alcohol in the past 12 months, then the edited variables corresponding to HLTH22 were assigned a code of 99, where 99 = LEGITIMATE SKIP. Thus, in the editing procedures for HLTH22, precedence was given to assigning codes of 91 or 93 ahead of assigning a code of 99.

The edited variables corresponding to HLTH22 retained missing values if HLTH22 had been skipped because of the following:

- respondents reported healthcare visits in the past 12 months but it was unknown whether respondents were past year alcohol users; or
- respondents reported past year alcohol use but it was unknown whether respondents had any healthcare visits in the past 12 months.

In addition, the last category in question HLTH22 allowed respondents to report that a doctor or other health professional did not discuss the respondents' alcohol use with them during any healthcare visits that they had in the past 12 months (HLTH22 = 95). The CAI logic did not allow respondents to choose this last response if they chose any of the preceding five response options for discussions about alcohol use, or vice versa. If this last response option was chosen in HLTH22, then the edited variable HPALCNOT corresponding to this response option retained a code of 95. The edited variables HPALCMT through HPALCTX were assigned a code of 99 (LEGITIMATE SKIP). If respondents chose one or more of the first five response options in HLTH22, then HPALCNOT was assigned a code of 99 (LEGITIMATE SKIP).

The edited variables corresponding to HLTH22 were examples of variables with the modified codes for "enter all that apply" questions that were described in Section 2.4.4. If respondents chose any of the responses for categories 1 through 5 from HLTH22, then these responses that were chosen were assigned a code of 1 (Yes). Other responses from the list that were not chosen were assigned a code of 2 (No [not entered]).

10.2.4.4 Talks about Illicit Drug Use

Respondents were eligible to be asked question HLTH23 (HPDRGTALK) if they had healthcare visits in the past 12 months and they reported past year use of any of the following: marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or methamphetamine. For cocaine, heroin, and methamphetamine, lifetime users who did not report past year use in the corresponding substance use sections were eligible to be asked HLTH23 if they reported past

year use of these substances in the special drugs section (e.g., use of cocaine with a needle in the past 12 months) (Section 7.2). When HLTH23 was skipped, HPDRGTALK was edited as described below.

- If question HLTH23 had been skipped and respondents never used any of the illicit drugs that were described in the preceding paragraph, then HPDRGTALK was assigned a code of 91, where 91 = NEVER USED MJ/COC/HER/HALL/INH/ METH. This code was assigned regardless of whether respondents reported any healthcare visits in the past 12 months.
- If question HLTH23 had been skipped because the most recent use of any of these illicit drugs was more than 12 months ago (which could include reports that respondents never used some of these types of illicit drugs), then HPDRGTALK was assigned a code of 93, where 93 = DID NOT USE MJ/COC/HER/HAL/INH/METH IN THE PAST 12 MONTHS. This code was assigned regardless of whether respondents reported any healthcare visits in the past 12 months.
- If question HLTH23 had been skipped because respondents had no healthcare visits in the past 12 months but they used any of these illicit drugs in the past 12 months, then HPDRGTALK was assigned a code of 99, where 99 = LEGITIMATE SKIP.
- Precedence was given to assigning codes of 91 or 93 ahead of assigning a code of 99.

HPDRGTALK retained missing values if HLTH23 had been skipped because of the following:

- respondents reported healthcare visits in the past 12 months but it was unknown whether respondents were users of any of these illicit drugs in the past year; or
- respondents reported past year use of one or more illicit drugs but it was unknown whether respondents had any healthcare visits in the past 12 months.

10.2.5 Specific Health Conditions

Respondents were asked about the following health conditions:

- sexually transmitted diseases in the past 12 months (question HLTH24), and
- the occurrence of the following diseases or conditions at any point in the respondent's lifetime (question HLTH25):
 - any kind of heart condition or heart disease;
 - diabetes or sugar diabetes;
 - chronic bronchitis, emphysema, or chronic obstructive pulmonary disease, also known as COPD;
 - cirrhosis of the liver;
 - hepatitis B or C;
 - kidney disease, not including bladder infection or incontinence;
 - asthma;
 - HIV or AIDS;
 - cancer or a malignancy of any kind; or
 - hypertension, also called high blood pressure.

Respondents who reported in HLTH25 that they ever had cancer were asked to choose from a list to report what kind of cancer(s) they had (question HLTH26), including other forms of cancer not on the list and to report how old they were when that type of cancer was first diagnosed (e.g., question HLTH28M for when their lung cancer was first diagnosed). Some of the cancers listed in HLTH26 were specific to males assigned at birth (prostate, testis) or to females assigned at birth (cervix, ovary, uterus), such that respondents were not shown the sex-specific cancers in HLTH26 that were assumed not to apply to them (e.g., females were not shown the response options for prostate or testicular cancer). Respondents who reported having some other kind of cancer were asked to specify the other kind of cancer that they had. Respondents who reported having any cancer in their lifetime also were asked if they had cancer in the past 12 months (question HLTH29).

Except for hypertension, respondents who reported that they had a specific condition in HLTH25 were asked how old they were when that particular condition was first diagnosed (e.g., question HLTH30 for when the respondent's heart disease was first diagnosed), or in the case of HIV or AIDS, how old they were when they first found out that they had HIV or AIDS. For hypertension, respondents first were asked whether they were currently taking medication for their high blood pressure (question HLTH40). Only those respondents who were currently taking medication for their high blood pressure were asked how old they were when their high blood pressure was first diagnosed (question HLTH41). If respondents reported that a particular condition was first diagnosed at an age that was greater than their current age, then the corresponding edited variable was assigned a code for bad data. Respondents who were diagnosed prior to age 1 were instructed to enter 1.

In addition, for heart disease, respondents who ever had heart disease were asked whether they had any kind of heart disease or a heart condition in the past 12 months (question HLTH31). Respondents who reported that they ever had asthma were asked in question HLTH38 whether they still had asthma.

The questions for the lifetime occurrence of specific health conditions in HLTH25 or specific cancers in HLTH26 were "enter all that apply" questions, in which respondents could choose more than one condition or more than one type of cancer from the list. Therefore, separate variables were created for each of the response choices for health conditions or specific cancers in HLTH25 and HLTH26 (e.g., CANCEVER for any lifetime cancer in HLTH25 and CALUNG for lung cancer in HLTH26). The corresponding edited variables were examples of variables with the modified codes for "enter all that apply" questions that were described in Section 2.4.4. If respondents chose any of the responses from HLTH25 or HLTH26, then these responses that were chosen were assigned a code of 1 (Yes). Other responses from the list that were not chosen were assigned a code of 2 (No [not entered]). Respondents also had the option to key an answer of "95" in HLTH25 (edited variable NONABOVEV) to indicate that they never had any of the health conditions listed in the question.

An important aspect of processing the variables for specific health conditions involved assignment of legitimate skip codes, where relevant (Section 2.4.2). Specific edits for assigning legitimate skip codes are described below.

- If NONABOVEV = 95 because respondents reported that they never had any of the conditions in HLTH25, then all remaining health condition variables were assigned legitimate skip codes.
- If a specific health condition was assigned a code of 2 (No [not entered]) but respondents reported other health conditions in HLTH25, then all related variables pertaining to that health condition were assigned legitimate skip codes. For example, if CANCEREVR was assigned a code of 2 (No [not entered]), then all variables pertaining to cancer were assigned legitimate skip codes.
- If CANCEREVR = 1 and respondents did not report having specific kinds of cancer in HLTH26, then the variables pertaining to these forms of cancer that respondents did not report having in their lifetime were assigned legitimate skip codes.
- If CANCEREVR = 1, variables for kinds of cancer that were specific to males (i.e., prostate, testicular) or to females (cervical, ovarian, uterine) typically were assigned legitimate skip codes according to the respondent's sex. For example, legitimate skip codes were assigned to variables pertaining to prostate cancer and testicular cancer if these forms of cancer were not reported and respondents were female. However, data were not overwritten with logically assigned legitimate skip codes or codes for bad data if respondents reported forms of cancer that did not agree with their current reported sex.
- If respondents reported that they were first diagnosed with cancer or with a heart condition at their current age, they were not asked whether they had the respective condition in the past year. The corresponding edited variables (CANCERYR and HRTCONDYR, respectively) were assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

A limited amount of "back editing" of the cancer variables also was conducted based on "OTHER, Specify" data for other forms of cancer (Section 2.4.5). A code of 3 (Yes LOGICALLY ASSIGNED) was assigned to a given cancer variable if respondents explicitly specified that they ever had that form of cancer. For example, if female respondents did not report having uterine cancer but explicitly specified uterine cancer as some other form of cancer that they had in their lifetime, then the edited variable CAUTERUS was assigned a code of 3 (Yes, LOGICALLY ASSIGNED). Because of the potential complexity in mapping other forms of cancer that respondents specified to specific organs or body systems, however, this editing was limited to situations in which "OTHER, Specify" codes were assigned that corresponded to the specific cancers listed in HLTH26. Nevertheless, analysts would have the option to use these "OTHER, Specify" codes in their analyses of NSDUH data for specific types of cancer.

10.3 Mental Health Services Utilization

Beginning in 2022, the section on mental health services utilization replaced sections for adult mental health service utilization and youth mental health service utilization. The new questionnaire section asked adult and youth respondents about the following: (1) receipt of specific sources of inpatient or outpatient mental health treatment in the past 12 months; (2) receipt of mental health treatment in the past 12 months using prescription medication, treatment via telehealth (i.e., over the phone or through video), or treatment in a prison, jail, or

juvenile detention center; (3) receipt of services in the past 12 months from other sources; (4) perceived unmet need for treatment; and (5) reasons for not receiving mental health treatment in the past 12 months.

10.3.1 Type and Location of Mental Health Treatment or Other Services

Sources of inpatient mental health treatment in the past 12 months that were asked about included (1) a hospital, (2) a residential mental health treatment center, (3) a residential drug or alcohol rehab or treatment center, and (4) some other place where respondents stayed overnight or longer. Sources of outpatient mental health treatment in the past 12 months that were asked about included (1) a mental health treatment center; (2) a drug or alcohol use treatment or rehab center; (3) the office of a therapist, psychologist, psychiatrist, or mental health professional; (4) a general medical clinic or doctor's office; (5) a hospital; (6) a school health or counseling center; and (7) some other place as an outpatient. Respondents also were asked whether they took medication in the past 12 months that was prescribed to them to help with their mental health, emotions, or behavior or whether they received mental health treatment in this period via telehealth (over the phone or through video) or in a prison, jail, or juvenile detention center. Sources of other services included (1) participation in a support group, (2) help from a peer support specialist or recovery coach, and (3) being seen in an emergency room or emergency department.

Respondents who indicated that they received treatment at some other place where they stayed overnight or longer were asked in "OTHER, Specify" question TXMHIPDO to describe the other place where they received treatment. Similarly, respondents who indicated that they received treatment at some other place as an outpatient were asked in "OTHER, Specify" question TXMHOPGO to describe the other place where they received outpatient treatment. Principles for coding these "OTHER, Specify" variables are discussed in Section 2.3.1.3. If respondents reported that they did not receive inpatient or outpatient treatment in "some other place," the respective inpatient or outpatient edited "OTHER, Specify" variable (MHTINOPLSP or MHTOUTOPLSP) was assigned a legitimate skip code.

Subsequent sections for locations where respondents received mental health treatment or other mental health services discuss edits according to respondents' answers in the "OTHER, Specify" questions. [Table B.21](#) discusses these edits in more detail.

10.3.1.1 Editing of Inpatient and Outpatient Treatment Variables

Consistent with the procedures described in Section 2.4.5, if respondents did not report that they received inpatient treatment in the past 12 months in a particular location in questions TXMJIPTA to TXMHIPTC, but their response to question TXMHIPDO was coded to indicate that respondents received inpatient treatment in the past 12 months in one of these previously mentioned locations, then the corresponding edited variable for that location was assigned a code of 3 (Yes LOGICALLY ASSIGNED). For example, if respondents did not report that they received inpatient treatment in a hospital, but they specified "hospital" as the other place where they stayed overnight to receive mental health treatment, then they were logically inferred to have received treatment in a hospital as an inpatient in the past 12 months. Similarly, if

respondents did not report that they received outpatient treatment in the past 12 months in a particular location in questions TXMHOPTA to TXMHOPTF, but they specified that location in question TXMHOPGO as the other place where they received outpatient treatment in the past 12 months, then the corresponding edited variable for that location was assigned a code of 3.

In addition, a small number of respondents in 2023 (fewer than five) reported that they received mental health treatment in some other place as an “outpatient” in the past 12 months, but they specified a residential location where they would be expected to stay overnight. For these respondents, the edited variable MHTINOPL for other inpatient locations was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTOUTOPLSP)] to indicate that inpatient treatment was received but respondents reported the inpatient location in an outpatient question. Similar edits were made for respondents who specified outpatient locations as other “inpatient” locations where they received mental health treatment, which applied to a small number of respondents in 2023 (fewer than five).

10.3.1.2 Editing of Variables for Treatment via Telehealth or in a Prison, Jail, or Juvenile Detention Center

Respondents could specify in questions TXMHIPDO or TXMHOPGO that the other location where they received “inpatient” or “outpatient” mental health treatment was via telehealth or in a prison, jail, or juvenile detention center. In these situations, the edited variables MHTPHVID (treatment by phone or video) or MHTPRISON (treatment in a prison, jail, or juvenile detention center) were assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP or MHTOUTOPLSP)].¹⁶⁰

10.3.1.3 Editing of Variables for Other Mental Health Services

Respondents could specify in questions TXMHIPDO or TXMHOPGO that the other location where they received “inpatient” or “outpatient” mental health treatment was in a support group, from a peer support specialist or recovery coach, or in an emergency room or emergency department. In these situations, the edited variables MHTSUPPGRP (support group), MHTPEERSPC (peer support specialist or recovery coach), or MHTERED (emergency room or department) were assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP or MHTOUTOPLSP)].

10.3.2 Perceived Unmet Need for Mental Health Treatment

The perceived unmet need for mental health treatment was measured in the 2023 NSDUH through the following sets of questions if respondents did not report that they received inpatient mental health treatment, outpatient mental health treatment, treatment via telehealth, or treatment in a prison, jail or juvenile detention center in the past 12 months or that they

¹⁶⁰ Because the inpatient and outpatient mental health treatment questions apply to locations where respondents received treatment, it is unlikely for respondents to specify taking prescription medication as “some other place” where they received inpatient or outpatient treatment. If this situation were to occur, the code of 5 would also be used to infer that respondents took medication that was prescribed to them to help with their mental health, emotions, or behavior.

took medication in that period that was prescribed to them to help with their mental health, emotions, or behavior:

- If respondents also did not report that they received services in a support group, from a peer support specialist or recovery coach, or in an emergency room or emergency department, they were asked one or both of the following questions:
 - These respondents first were asked whether they sought professional counseling, medication, or other treatment for their mental health, emotions, or behavior (question TXMHSEK).
 - Respondents who did not report in TXMHSEK that they sought counseling, medication, or other treatment were asked whether they thought they should get professional counseling, medication, or other treatment for their mental health, emotions, or behavior (question TXMHGET).
- If respondents reported that they received services in a support group, from a peer support specialist or recovery coach, or in an emergency room or emergency department, they were asked one or both of the following questions:
 - These respondents first were asked whether they sought additional professional counseling, medication, or other treatment for their mental health, emotions, or behavior (question TXMHSEK2).
 - Respondents who did not report in TXMHSEK2 that they sought additional counseling, medication, or other treatment were asked whether they thought they should get additional professional counseling, medication, or other treatment for their mental health, emotions, or behavior (question TXMHGET2).

An important aspect of processing the variables for the perceived unmet need for mental health treatment involved assignment of legitimate skip codes, where relevant (Section 2.4.2), according to how respondents answered prior questions in the mental health services utilization section. For example, if respondents reported that they received inpatient treatment, outpatient treatment, treatment via telehealth, treatment in a prison, jail, or juvenile detention center or that they took prescribed medication for their mental health, then all variables pertaining to the perceived unmet need for mental health treatment were assigned legitimate skip codes.

[Table B.21](#) discusses specific edits pertaining to the variables for whether respondents sought treatment or thought they should get mental health treatment.

10.3.3 Reasons for Not Receiving Mental Health Treatment

Respondents who reported an unmet need for mental health treatment in the past 12 months (i.e., they sought or thought they should get mental health treatment or additional mental health treatment) were asked a series of questions regarding reasons for them not receiving treatment in that period. For each reason, respondents were asked whether that was one of the reasons or not one of the reasons for them not getting professional counseling, medication, or other treatment for their mental health (or for them not getting additional professional counseling, medication, or other treatment for their mental health).

Respondents were assigned a legitimate skip code of 99 for these questions in the following situations:

- They reported receiving inpatient or outpatient treatment, treatment via telehealth, treatment in a prison, jail, or juvenile detention center or taking prescribed medication.
- They were asked whether they sought treatment or thought they should get treatment (or additional treatment) for their mental health, but they reported that they did not seek treatment and they did not think that they should get it.

10.4 Adult Mental Health

The mental health section for adults included administration of the Kessler-6 (K6) distress scale to collect data on distress in the past 30 days and in the past 12 months (see Section B.4.5 in the 2008 national findings report for details; Office of Applied Studies, 2009). Respondents were routed into the 12-month version of the K6 if they reported having a period in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt in the past 30 days. Impairment associated with psychological distress has been assessed using an abbreviated World Health Organization Disability Assessment Schedule (WHODAS) (see Section B.4.6 in the 2008 national findings report for details; Office of Applied Studies, 2009). Questions about suicidal thoughts and behavior in the past 12 months followed the questions on impairment.

10.5 Adult and Adolescent Depression

Questions in adult and adolescent depression sections were based on those used in the [National Comorbidity Survey Replication](#) [↗](#) for adults (NCS-R) and the [National Comorbidity Survey Replication Adolescent Supplement](#) [↗](#) (NCS-A). These depression sections were included in NSDUH to produce lifetime and 12-month prevalence estimates of major depressive episode (MDE), severity of 12-month MDE, age at first MDE, lifetime number of episodes, current and 12-month treatment, and the respondent's perception of treatment effectiveness.

There were some differences in wording between related items in these sections, such as use of simpler wordings for the adolescent depression questions. For example, question AD17 in the adult depression section asked respondents how "severe" their "emotional distress" was during their worst periods lasting 2 weeks or longer when they had problems with their mood. The corresponding question YD17 from the adolescent depression section asked adolescents how "strong" their "bad feelings" were during these periods. Despite these differences in wording, similar naming conventions were used for the variables in these two sections. For example, the edited variable corresponding to question AD17 from the adult depression section was ADWRDST (where WR = worst period, and DST = distress), and the edited variable corresponding to question YD17 from the adolescent depression section was YOWRDST, even though YD17 did not ask about distress. Thus, the only difference in the names for analogous edited variables in these sections was in the use of the two-letter prefix that defined which section a given variable came from: "AD" for variables from the adult depression section and "YO" for adolescent depression variables.

There also were differences in how the CAI program created indicators of MDE for adolescents and adults based on differences in the NCS-A and NCS-R. Specifically, the criteria for defining respondents as having the symptom of loss of interest or pleasure in most things was less restrictive for adolescents than for adults. In particular, the *Diagnostic and Statistical Manual of Mental Disorders*, 4th edition (DSM-IV; American Psychiatric Association, 1994), criteria for MDE place more emphasis for adolescents on the cognitive aspects of depression, such as boredom or apathy, rather than on somatic or physical complaints, such as sleep loss, that may be manifest in adults with MDE. For example, somatic or physical complaints, such as sleep loss, may be due to factors during adolescence other than depression. Consequently, somatic or physical complaints that may be associated with MDE among adults function less well as indicators of MDE among adolescents than do cognitive indicators. For this reason, the CAI logic gave adolescents additional opportunities to be classified as having the symptom of loss of interest based on their answers to questions that were not considered in classifying adults as having this symptom.

Despite these differences between the adult and adolescent depression sections, the basic logic for asking questions was similar between the two sections. Therefore, except where specific differences are discussed in terms of how variables were edited for these sections, the same basic edits discussed below applied to variables in both sections.

An important aspect of the processing of variables in these sections consisted of assigning legitimate codes based on the routing logic within these sections. In particular, adults were assigned legitimate skip codes to the edited variables in the adolescent depression section, and adolescents were assigned legitimate skip codes to the edited variables in the adult depression section.

As an additional example, respondents were asked a series of questions to identify changes in appetite or weight. They first were asked whether they had a much smaller appetite than usual during the most recent time when their problems were the worst (questions AD26A and YD26A, corresponding to edited variables ADWRELES and YOWRELES, respectively). If respondents answered the relevant question as “yes,” they were skipped out of subsequent questions about increases in appetite and weight gain. If respondents reported having less appetite, they were asked whether they lost weight without trying to, and if so, whether their weight loss was due to them being sick or on a diet. Respondents who indicated that their weight loss was due to sickness or a diet were not asked to report how many pounds they lost in weight.

Conversely, if respondents did not report in AD26A/YD26A that they had less appetite than usual, they were asked whether they had a much larger appetite than usual (questions AD26B and YD26B, corresponding to edited variables ADWREMOR and YOWREMOR, respectively). Respondents who reported that they had a much larger appetite were skipped out of remaining questions related to weight loss. These respondents subsequently were asked whether they gained weight without trying to. If respondents reported gaining weight without trying to, the CAI program asked follow-up questions to rule out weight gains due to growth (for respondents aged 21 or younger) or pregnancy (for females). Respondents who indicated that they gained weight because they were growing or because they were pregnant were not asked to report

how many pounds they gained. Thus, editing of the adult and adolescent depression variables related to changes in appetite or weight involved assignment of legitimate skip codes to these variables based on the routing logic for the corresponding questions.

As noted in Section 2.4.3, legitimate skip codes generally were not assigned if a lead question was answered as “don’t know” or “refused.” However, important exceptions to this principle were made in editing of the adult and adolescent depression variables because of consideration of other aspects of the routing logic in these sections. In particular, the lead screening questions for adult depression (ASC21,¹⁶¹ ASC22,¹⁶² and ASC23,¹⁶³ corresponding to edited variables ADDPREV, ADDSCEV, and ADLOSEV) and the lead screening questions for adolescent depression (YDS21,¹⁶⁴ YDS22,¹⁶⁵ and YDS23,¹⁶⁶ corresponding to edited variables YODPREV, YODSCEV, and YOLOSEV) at the beginning of these sections had a special skip logic. Respondents were routed through the three screening questions until either they gave an affirmative answer to one of the questions or all screening questions had been asked. An affirmative answer to any of the lead questions routed respondents into follow-up questions that could screen respondents into further questions about depression.

In the adult depression section, for example, if question ASC21 was answered as “don’t know” or “refused,” the subsequent question AD01¹⁶⁷ was skipped. In this situation, however, respondents were routed to follow-up question ASC22. Thus, if respondents answered question ASC21 “don’t know” or “refused” but they answered question ASC22 as “yes,” they were still eligible to be administered the remainder of the adult depression section, depending on how they answered subsequent questions.

Therefore, legitimate skip codes were assigned (where relevant) to the variables corresponding to questions AD01 through AD09 in the adult depression section (edited variables ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK) if at least one item from questions ASC21, ASC22, or ASC23 was answered as “yes” or “no.” For example, if question ASC21 was answered as “don’t know,” and question ASC22 or ASC23 was answered as “yes” or “no,” the editing procedures assigned a legitimate skip code to ADDPDISC, corresponding to question AD01. Similarly, if

¹⁶¹ Question ASC21 asked, “Have you ever in your life had a period of time lasting several days or longer when most of the day you felt sad, empty, or depressed?”

¹⁶² Question ASC22 asked, “Have you ever had a period of time lasting several days or longer when most of the day you were very discouraged about how things were going in your life?”

¹⁶³ Question ASC23 asked “Have you ever had a period of time lasting several days or longer when you lost interest in most things you usually enjoy like work, hobbies, and personal relationships?”

¹⁶⁴ Question YDS21 asked, “Have you ever in your life had a period of time lasting several days or longer when most of the day you felt sad, empty, or depressed?”

¹⁶⁵ Question YDS22 asked, “Have you ever had a period of time lasting several days or longer when most of the day you felt very discouraged or hopeless about how things were going in your life?”

¹⁶⁶ Question YDS23 asked “Have you ever had a period of time lasting several days or longer when you lost interest and became bored with most things you usually enjoy like work, hobbies, and personal relationships?”

¹⁶⁷ If respondents answered question ASC21 as “yes,” they were asked question AD01. Question AD01 asked, “During times when you felt sad, empty, or depressed most of the day, did you ever feel discouraged about how things were going in your life?”

question ASC22 was answered as “don’t know” but ASC23 was answered as “yes” or “no,” the editing procedures assigned a legitimate skip code to ADDSLSIN, corresponding to question AD02. In turn, if ADDPDISC, ADDPLSIN, and ADDSLSIN all were answered as “no” or had legitimate skip codes after the above edits, then ADLSI2WK (corresponding to AD09) was assigned a legitimate skip code. The values in ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK determined whether subsequent variables were assigned legitimate skip codes. The logic provided in this example for adult depression also was applied to the corresponding variables in the adolescent depression section.

As discussed in Section 2.4.3, if a lead question that governs a skip pattern was refused, the editing procedures typically “propagated” that refusal from the lead question to the variables that had been skipped. For most of the adult and adolescent depression variables, however, this refusal propagation was not performed. The CAI program contained routines for scoring the symptom indicators for MDE. The CAI program coded the symptom score variable DSMMDEA2 as 2 (i.e., has fewer than 5 symptoms) if the sum of the numbers of codes of 1 (i.e., has symptom), “don’t know,” or “refused” in the individual symptom indicators was less than 5. That is, even if the unknown responses (“don’t know” or “refused”) had instead been coded as 1 to indicate the presence of a symptom, respondents would not have reached the threshold of 5 symptoms for DSMMDEA2 to be coded as 1. Therefore, not propagating refusals helped to avoid situations in which a different overall score might be obtained if analysts were to calculate DSMMDEA2 based on edited variables.

The exception to this rule of not propagating refusals in the adult and adolescent depression sections concerns the final questions regarding receipt of counseling from a medical doctor or other professional about the respondents’ symptoms of depression (e.g., questions AD86 through AD86F in the adult depression section). For example, if question AD86 (edited variable ADSEEDOC) was refused, the editing procedures still propagated that refusal code to the skipped variables that were dependent on AD86.

In addition, the CAI program created MDE symptom variables and overall MDE symptom scores for adults and adolescents. [Table 10.1](#) lists the final, edited variables that were created from these symptom variables and overall symptom scores. For each variable, explanation of the meaning of that variable also is provided. The only editing that was done to these variables in [Table 10.1](#) was to assign legitimate skip codes based on the respondent’s age (i.e., 12 to 17 or 18 or older). Thus, the values that were created by the CAI program were preserved in the edited symptom variables.

Table 10.1 Depression Symptom and Score Variables

Adult Depression Variable	Adolescent Depression Variable	Explanation
AD_MDEA1	YO_MDEA1	Respondent (R) felt sad, empty, depressed, or discouraged most of the day.
AD_MDEA2	YO_MDEA2	R lost interest or pleasure in most things.
AD_MDEA3	YO_MDEA3	R had changes in appetite or weight (not due to growth, pregnancy, illness, or dieting).

Table 10.1 Depression Symptom and Score Variables (continued)

Adult Depression Variable	Adolescent Depression Variable	Explanation
AD_MDEA4	YO_MDEA4	R had sleep problems.
AD_MDEA5	YO_MDEA5	Others noticed that the R was restless or lethargic.
AD_MDEA6	YO_MDEA6	R felt tired or low on energy nearly every day.
AD_MDEA7	YO_MDEA7	R felt worthless nearly every day.
AD_MDEA8	YO_MDEA8	R was unable to concentrate or make decisions.
AD_MDEA9	YO_MDEA9	R was suicidal (had thoughts of suicide, made plans, or made an attempt).
ADSMMDA	YODSMMDA	Score of symptom indicators 1 through 9 from above.

Relatively little additional editing was done to the adult and adolescent depression variables, aside from assigning legitimate skip codes. Additional editing issues that were relevant to these sections are described in the remainder of this section.

If respondents reported a period of time when their symptoms or problems were the worst, they were asked to report how old they were when this time started (edited variables ADWRAGE for adults and YOWRAGE for adolescents). In addition, if respondents scored as positive for MDE (edited variables ADSMMDEA and YODSMMDA) and they reported that these problems caused some interference with their work, social life, or relationships, they were asked to report the age at which these problems first occurred (edited variables ADPBAGE and YOPBAGE). If respondents reported an age of onset in any of these variables that was greater than their current age, these variables were set to bad data. For adults, if ADWRAGE had been set to bad data and the respondent's original answer was age 22 or greater, the respondent was skipped out of the question that asked about weight gain because the respondent was growing (edited variable ADWRGROW). Therefore, if ADWRAGE had been set to bad data, ADWRGROW retained a code of blank.

As noted previously, if respondents reported gaining or losing weight, and these gains or losses could not be attributed to factors other than depression (e.g., growth, pregnancy, dieting), respondents were asked to report the number of pounds they gained or lost. Respondents were allowed to report gains or losses of pounds only in terms of whole numbers and could report that they gained or lost 0 pounds. No editing was done to the variables ADWRGNLB, ADWRSLB, YOWRGNLB, or YOWRSLB when the response of 0 pounds occurred, because a response of 0 could refer to a gain or loss of a fraction of a pound. Furthermore, only gains or losses of 10 or more pounds resulted in respondents being scored as having the symptom of changes in appetite or weight.

In addition, a feature of the logic for asking respondents about gains or losses in weight was that if respondents were asked the question about gaining weight without trying and they answered it as "don't know" or "refused," they had an additional opportunity to be asked questions about losing weight without trying. That is, the program was looking for the first

affirmative set of answers that would allow a determination to be made of whether respondents gained or lost enough pounds to qualify for being depressed. Consequently, no editing was done if respondents originally gave an answer in the questions corresponding to ADWREMOR or YOWREMOR (i.e., having a much larger appetite than usual), they answered the weight gain question (corresponding to ADWRGAIN or YOWRGAIN) as “don’t know” or “refused,” and then they were routed into the questions about weight loss.

If respondents reported that they talked to a medical doctor or other professional in the past 12 months about the problems they were experiencing related to depression, they were asked to report which professionals they saw or talked to. In the question pertaining to the specific professionals that they saw or talked to, respondents were allowed to enter more than one type of professional from the list they were presented. As in other sections, the “enter all that apply” variables in the adult and adolescent depression sections were coded as 1 if the response was entered and as 6 if the response was not entered (Section 2.4.4).

Codes of 94 and 97 (for “don’t know” and “refused,” respectively) were assigned to an entire list of variables if respondents did not know or refused to report the specific professionals that they saw or talked to about their problems. If the entire list was blank but respondents had previously reported that they saw or talked to a professional about their problems, then the specific variables corresponding to categories of helping professionals retained a code of 98 (blank).

Respondents could report that they saw or talked to “another type of helping professional” and then specify a helping professional that they had already been asked about, such as a psychiatrist. Thus, for example, if the edited variable ADPSYMD, pertaining to services from a psychiatrist, was not coded as 1 and respondents specified that they saw or talked to a psychiatrist, ADPSYMD was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED.

10.6 COVID-19

Beginning in Quarter 4 of 2020, a section was included in the questionnaire on the perceived impact of the COVID-19 pandemic on respondents’ mental health, substance use, finances, living situation, and access to services. This section applied to adults and youths. As noted in Section 2.3.2.2, responses in the COVID-19 section were replaced with bad data codes when patterned responses were identified.

Consistent with the general principle of not editing across sections of the interview (Section 2.4.1), data from these COVID-19 items were not edited for consistency with data from other sections (or vice versa), except when responses to questions in earlier sections governed whether respondents received questions in the COVID-19 section. For example, any inconsistencies between responses to questions regarding service use in the COVID-19 section and questions regarding substance use treatment or mental health service use in previous questionnaire sections were not resolved through editing.

However, data are edited across sections when responses in one or more earlier sections governed whether respondents were asked questions in a later section. In the COVID-19 section, only respondents who reported that they used alcohol in the past 12 months were asked question COV2 regarding changes in their alcohol use since the COVID-19 pandemic began. Similarly, only respondents who reported the use of drugs other than alcohol¹⁶⁸ in the past 12 months were asked question COV3 regarding changes in their use of other drugs since the COVID-19 pandemic began. Therefore, editing of COV2 and COV3 took respondents' prior answers into account for whether respondents used alcohol or other drugs in the past 12 months. The editing procedures for COV2 and COV3 are described in further detail in Sections 10.6.1 and 10.6.2.

No editing was done to the remaining edited variables in the COVID-19 section other than to assign descriptive names.

10.6.1 Editing for Skip Logic on the Perceived Effect of the COVID-19 Pandemic on Alcohol Use

Only respondents who reported use of alcohol within the past 12 months in the alcohol section of the questionnaire were eligible to be asked question COV2 regarding how the COVID-19 pandemic affected respondents' level of alcohol use. Data on the lifetime and the most recent use of alcohol use data were used to edit COV2 (edited variable COALCUSE).

- If question COV2 was skipped and respondents *never* used alcohol, then COALCUSE was assigned a code of 91 (NEVER USED ALCOHOL).
- If question COV2 was skipped and respondents last used alcohol more than 12 months prior to the interview, then COALCUSE was assigned a code of 93 (DID NOT USE ALCOHOL IN THE PAST 12 MONTHS).
- If question COV2 was skipped because it was unknown whether respondents ever used alcohol, or they used alcohol in their lifetime, but the period of their most recent alcohol use was unknown, then COALCUSE retained a code of 98 (BLANK).
- If question COV2 was asked but respondents' alcohol use data had been set to bad data, then COALCUSE was assigned a code of 85 (BAD DATA Logically Assigned).

10.6.2 Skip Logic for Effect of COVID-19 Pandemic on Other Drug Use

Similar to COV2, only respondents who reported use of marijuana, cocaine, crack, heroin, hallucinogens, inhalants, or methamphetamine in the past 12 months, or *any* use of pain relievers, tranquilizers, stimulants, or sedatives (i.e., not necessarily misuse) in that period, were eligible to be asked question COV3 regarding how the COVID-19 pandemic affected respondents' level of use of drugs other than alcohol. The following describes how edited variables corresponding to recency of drug use were used to edit CODRGUSE.

¹⁶⁸ Use of drugs other than alcohol included the use of marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or methamphetamine in the past year or any use (i.e., not necessarily misuse) of prescription pain relievers, tranquilizers, stimulants, or sedatives in the past year.

- If question COV3 was skipped and respondents never used any of these drugs, then CODRGUSE was assigned a code of 91 (NEVER USED DRUGS).
- If question COV3 was skipped because respondents did not use any of these drugs in the past 12 months and the most recent use of at least one of these drugs was more than 12 months ago (which could include reports that respondents never used some of these drugs), then CODRGUSE was assigned a code of 93 (DID NOT USE DRUGS IN THE PAST 12 MONTHS).
- If question COV3 was skipped because lifetime use or the period of last use of all of these drugs was unknown, then CODRGUSE retained a code of 98 (BLANK).
- If question COV3 was asked but all drug use variables responsible for COV3 being asked were set to bad data, then CODRGUSE was assigned a code of 85 (BAD DATA Logically Assigned).¹⁶⁹

10.7 Imputation

Missing data for select mental health services utilization, adult mental health, and adult depression variables were imputed. Section 10.7.1 covers the imputation of the mental health services utilization variables. Section 10.7.2 covers the imputation of the adult mental health variables and adult depression variables.

10.7.1 Mental Health Services Utilization

Missing data for variables related to mental health services utilization were imputed using the modified predictive mean neighborhood (modPMN) methodology as described in Chapter 3. Section 10.7.1.1 describes how these variables were segregated into imputation sets, and Section 10.7.1.2 provides details on the steps of modPMN and the sequencing of steps.

10.7.1.1 Imputation Sets

The mental health services utilization variables were divided into two imputation sets as shown in [Table 10.2](#). These imputation sets are independent of each other and were run in parallel. As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is by descending item response rates and can change from year to year. [Table 10.2](#) lists the final imputation-revised mental health services utilization variables in each imputation set. An imputation indicator was delivered with each imputation-revised variable. Similar to the naming convention of other imputation indicators, the prefix IR of the imputation-revised variable is replaced with II.

¹⁶⁹ For example, if respondents were routed to COV3 based only on reported marijuana use in the past 12 months (i.e., all other drugs were either never used or used more than 12 months ago), but the marijuana data had been set to bad data, then CODRGUSE was set to 85.

Table 10.2 Mental Health Services Utilization Variables in Imputation Sets 1 and 2

Variable Description	Final Imputed Variable	Imputation Set
Received Inpatient Mental Health Treatment		
Hospital	IRMHTINHOSP	1
Residential Drug or Alcohol Rehab or Treatment Center	IRMHTINRHAB	1
Residential Mental Health Treatment Center	IRMHTINRSMH	1
Some Other Place as an Inpatient	IRMHTINOPL	1
Received Outpatient Mental Health Treatment		
Drug or Alcohol Rehab or Treatment Center	IRMHTOUTRHAB	1
Mental Health Treatment Center	IRMHTOUTMHCR	1
Office of a Therapist, Psychologist, Psychiatrist, or Substance Use Treatment Professional	IRMHTOUTTHRP	1
General Medical Clinic or Doctor's Office	IRMHTOUTDOC	1
Hospital	IRMHTOUTHOSP	1
School Health or Counseling Center	IRMHTOUTSCHL	1
Some Other Place as an Outpatient	IRMHTOUTOPL	1
Received Other Treatment or Services for Mental Health		
Prescription Medication	IRMHTRXMED	1
Telehealth Treatment	IRMHTPHVID	1
Prison, Jail, or Juvenile Detention Center	IRMHTPRISON	1
Support Group	IRMHTSUPPGRP	2
Emergency Room/Emergency Department	IRMHTERED	2
Peer Support Specialist or Recovery Coach	IRMHTPEERSPC	2

10.7.1.2 Imputation Process

The mental health services utilization variables are all dichotomous. Imputation of categorical variables using the modPMN method is described in Sections 3.6.1 through 3.6.3.

10.7.1.2.1 Response Propensity Step

The overall process for the response propensity step is described in Section 3.6.1. For the mental health services utilization variables, specifically, two imputation cycles were run for each set. This required fitting 30 response propensity models for set 1 and 8 models for set 2. The following are the four key inputs to the response propensity adjustment step:

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain included all respondents.
- **Item Response Indicator:** If a valid response was provided for the given outcome, the individual was deemed an item respondent.

- **Covariates:** See Appendix D for information on the covariates used in the response propensity models.

10.7.1.2.2 Prediction Modeling Step

The SUDAAN® LOGISTIC procedure (called RLOGIST in SAS®-callable SUDAAN) was used to model dichotomous variables like the mental health services utilization variables.¹⁷⁰ All mental health services utilization variables had the same levels for the outcome variable:

- received the specified treatment or service in the past 12 months, or
- did not receive the specified treatment or service in the past 12 months.

10.7.1.2.3 Collocated Stochastic Imputation Step

This section provides specific information about the collocated stochastic imputation step for the mental health services utilization variables. Section 3.6.3 provides general information about the collocated stochastic imputation step. Each of the prediction models is followed by an imputation step where the predicted probabilities from the previous step, along with a random number, is used to impute whether the item nonrespondent received the specified treatment or service. Partial information is not available for these edited variables. Therefore, no adjustments to the predictive mean vectors are made.

10.7.2 Adult Mental Health and Adult Depression Variables

Missing data for the adult mental health variables from the K6 distress scale and the WHODAS were statistically imputed beginning in 2021. Adult depression variables related to MDE and suicidal thoughts and behaviors were also statistically imputed. For these variables, the modified predictive mean neighborhood (modPMN) methodology was used to replace missing values as described in Chapter 3. Section 10.7.2.1 describes how these variables were segregated into imputation sets, Section 10.7.2.2 describes how the variables were ordered within the imputation sets, and Section 10.7.2.3 provides details on the modPMN steps and the sequencing of steps.

10.7.2.1 Imputation Sets

The adult mental health and adult depression variables imputed using the modPMN methodology are segregated into 9 imputation sets as outlined in [Table 10.3](#). Only respondents who indicate during the administration of the K6 items that they had experienced at least one of the K6 symptoms at least a “little of the time” are administered the WHODAS. Therefore, set 1 must be completed prior to set 2 to allow the domain for the WHODAS imputations to be fully defined. Because past year MDE is restricted to adults with a lifetime MDE, sets 3, 4, and 5 must be completed sequentially. Likewise, set 6 is used to define the domain for sets 7 and 8 (e.g., only respondents who reported a suicide attempt in the past year were asked whether this was due to COVID-19 and whether they received medical attention for their suicide

¹⁷⁰ SUDAAN® is a registered trademark of Research Triangle Institute. SAS® software is a registered trademark of SAS Institute Inc.

attempt). Finally, only respondents who reported receiving medical attention for a suicide attempt in the past year were asked the question about whether they stayed overnight in a hospital for a suicide attempt, so set 8 is used to define the domain for set 9.

Table 10.3 Imputation Sets

Outcome Group	Imputation Set
Kessler-6, Past Month	1
Kessler-6, Month in Past Year Worse than Past 30 Days	
Kessler-6, Past Year	
World Health Organization Disability Assessment Schedule	2
Adult Lifetime MDE	3
Adult Past Year MDE	4
Adult Past Year MDE with Impairment	5
Adult Suicidal Thoughts in Past Year	6
Adult Suicide Plans in Past Year	
Adult Suicide Attempts in Past Year	
Adult Suicidal Thoughts in Past Year due to COVID-19	7
Adult Suicide Plans in Past Year due to COVID-19	
Adult Suicide Attempts in Past Year due to COVID-19	
Adult Received Medical Attention for Suicide Attempt in Past Year	8
Adult Stayed Overnight in Hospital for Suicide Attempt in Past Year	9

COVID-19 = coronavirus disease 2019; MDE = major depressive episode.

10.7.2.2 Imputation Order within Imputation Sets

As described in Section 3.7.1, the sequencing of variables in modPMN within an imputation set is generally by descending item response rates. However, because of the relationships between variables in the multiple-variable sets 1, 6, and 7, the imputation order within these sets is static and will not change from year to year. In set 1, the past year K6 questions only apply to respondents who reported that there was a month in the past year when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days. Therefore, the variable related to whether there was a worst month in the past year must be imputed prior to the past year K6 variables. The imputation order within sets 6 and 7 is also static because most adults who report suicide plans or attempts also report serious thoughts of suicide. Making the order static ensures that serious thoughts of suicide can be used as a covariate for suicide plans and suicide attempts in both cycles of modPMN. The ordering for sets 1, 6, and 7 is shown in [Table 10.4](#). Sets 2, 3, 4, 5, 8, and 9 are not included in [Table 10.4](#) because these sets contain only a single outcome.

Table 10.4 Sequencing of Outcomes within Multiple-Variable Imputation Sets

Imputation Set	Order
1: Kessler-6	Past Month, Month in Past Year Worse than Past 30 Days, Past Year
6: Adult Suicidal Thoughts and Behaviors	Thoughts, Plans, Attempts
7: Adult Suicidal Thoughts and Behaviors due to COVID-19	Thoughts, Plans, Attempts

COVID-19 = coronavirus disease 2019.

10.7.2.3 Imputation Process

The imputation procedures used to impute variables related to adult mental health and adult depression are described in detail in Section 3.6. Deviations from the processes described in these sections are described below.

10.7.2.3.1 Response Propensity Step

The overall process for the response propensity step was similar to the processes described in Section 3.6.1. However, the number of models fit and the inputs to the response propensity step varied from set to set. The total number of response propensity models fit for each set along with the domain can be found in [Table 10.5](#). Following are the four key inputs to the response propensity adjustment step.

- **Analysis Weight:** All sets used the preliminary analysis weight.
- **Domain:** The domain for each set is detailed in [Table 10.5](#).
- **Item Response Indicator:** If a valid response was provided for the given outcome, the individual was deemed an item respondent. For the past month K6, past year K6, and WHODAS imputations, an individual must provide a valid response for each item included in the scale to be deemed an item respondent.
- **Covariates:** See Appendix D for information on the covariates used in the response propensity models.

Table 10.5 Summary of Response Propensity Models

Imputation Set	Number of Response Propensity Models	Domain
1: Kessler-6	3	All adult respondents (past month Kessler-6; month in past year worse than past 30 days) Adult respondents with a month in the past year that was worse than past 30 days (past year Kessler-6)
2: World Health Organization Disability Assessment Schedule	1	Adult respondents who reported experiencing at least one of the Kessler-6 symptoms at least a “little of the time”
3: Adult Lifetime MDE	1	All adult respondents
4: Adult Past Year MDE	1	Adult respondents with lifetime MDE
5: Adult Past Year MDE with Impairment	1	Adult respondents with past year MDE

Table 10.5 Summary of Response Propensity Models (continued)

Imputation Set	Number of Response Propensity Models	Domain
6: Adult Suicidal Thoughts and Behaviors	3	All adult respondents
7: Adult Suicidal Thoughts and Behaviors due to COVID-19	3	Adult respondents with suicide thoughts, plans, attempts in past year
8: Medical Attention for Suicide Attempt	1	Adult respondents with a suicide attempt in past year
9: Hospital Stay for Suicide Attempt	1	Adult respondents who received medical attention for a suicide attempt in past year

COVID-19 = coronavirus disease 2019; MDE = major depressive episode.

10.7.2.3.2 Prediction Modeling Step

The outcomes in sets 3 through 9, as well as the worse month outcome in set 1, are dichotomous (i.e., Yes or No) and used dichotomous logistic regression through SUDAAN's LOGISTIC procedure (called RLOGIST in SAS-callable SUDAAN).

Two 6-item K6 scales are used for NSDUH to gather information regarding how frequently a respondent experienced symptoms of psychological distress. The two scales measure the same symptoms but during two different periods of time: during the past 30 days and during one month in the past 12 months when the respondent's symptoms were worse than in the past 30 days. For each K6 scale (i.e., past month and past year), six questionnaire items are coded with values ranging from 0 ("None of the time") to 4 ("All of the time"). For each time period, the six items are summed to create an overall score (KSSLR6MONED, KSSLR6YRED) ranging from 0 to 24, which is then used in the prediction modeling step as shown in [Table 10.6](#). Because the total K6 scores each contain 25 levels (i.e., values of 0 to 24), the outcome was treated as a continuous variable. Similar to the approach taken for modeling frequency of substance use, the empirical logit transformation was applied to the total scores prior to modeling. The empirical logit transformation was applied to the total scores as shown below, where $K_{i,t}$ is the total K6 score for respondent i in time t (i.e., past month or past year) and 24 is the maximum total score.

$$y_{i,t} = \log\left(\frac{K_{i,t} + 0.5}{24 - K_{i,t} + 0.5}\right)$$

The transformed variable, $y_{i,t}$, was then used as the dependent variable in a linear regression (i.e., the REGRESS procedure in SUDAAN) for each time period t and predicted values were obtained for both respondents and nonrespondents. Information on the covariates included in the linear regression model are provided in Appendix D.

Table 10.6 Edited Kessler-6 Variables

Reference Period	Edited Variable for Prediction Modeling	Edited Source Variable Name	Edited Source Variable Description
Past Month	KSSLR6MONED	DSTCHR30	During the past 30 days, how often felt so sad or depressed that nothing could cheer you up?
		DSTEFF30	During the past 30 days, how often did you feel that everything was an effort?
		DSTHOP30	During the past 30 days, how often did you feel hopeless?
		DSTNGD30	During the past 30 days, how often did you feel down on yourself, no good, or worthless?
		DSTNRV30	During the past 30 days, how often did you feel nervous?
		DSTRST30	During the past 30 days, how often did you feel restless or fidgety?
Worst Month in Past Year	KSSLR6YRED	DSTCHR12	During the worst month, how often felt so sad or depressed that nothing could cheer you up?
		DSTEFF12	During the worst month, how often did you feel that everything was an effort?
		DSTHOP12	During the worst month, how often did you feel hopeless?
		DSTNGD12	During the worst month, how often did you feel down on yourself, no good, or worthless?
		DSTNRV12	During the worst month, how often did you feel nervous?
		DSTRST12	During the worst month, how often did you feel restless or fidgety?

The WHODAS is a scale used to measure functional impairment that consists of a series of items that are used for assessing disturbances in social adjustment and behavior (i.e., functional impairment). A reduced set of 13 WHODAS items are included in the NSDUH as shown in [Table 10.7](#). These 13 questions ask respondents how much difficulty they experienced in doing eight daily activities in the one month in the past year when they were at their worst emotionally. Responses to this impairment scale are then transformed and recoded to create different WHODAS scores used during analysis.

The first WHODAS measure (WHODASSCED) codes each of the eight items on a scale of 0 ("No difficulty") to 3 ("Severe difficulty"). Some items contain a fifth category corresponding to responses of "Not applicable." If this category is selected, then a follow-up question asks the respondent whether emotions, nerves, or mental health prevented them from performing the activity. A response of "Yes" is coded as 3 and a response of "No" is coded as 0. After recoding, the eight individual items are then summed to create a total WHODAS score that ranges in value from 0 to 24.

The second WHODAS measure (WHODASDAED) is created to indicate the number of daily activities in which a respondent had moderate or severe difficulty performing or did not perform

due to problems with emotions, nerves, or mental health. Each of the eight WHODAS items are transformed into values of 0 or 1, where responses indicating “Moderate difficulty” or “Severe difficulty” are coded as 1 and responses of “Mild difficulty” or “No difficulty” are coded as 0. If the “Not applicable” response is available and selected, then a response of “Yes” in the follow-up question is coded as 1 and a response of “No” in the follow-up question is coded as 0. After transforming the variables, the eight items are summed to create the alternative WHODAS score that ranges from 0 to 8.

For the WHODAS imputations, two prediction models corresponding to the total score and the alternative score were used. As with the K6 prediction models, the empirical logit transformation was applied to the total score prior to modeling. The transformed total score was then used as the dependent variable in a linear regression model (i.e., the REGRESS procedure in SUDAAN). A second prediction model was also fit using the alternative score as the dependent variable. For this second model, a cumulative logit model (i.e., the MULTILOG procedure in SUDAAN with CUMLOGIT link function) was used. From each of these two models, predicted values were obtained for both respondents and nonrespondents. See Appendix D for information on the covariates included in the prediction models.

Table 10.7 Edited WHODAS Variables

Edited Variables for Prediction Modeling	Edited Source Variable Name	Edited Source Variable Description
Total Score: WHODASSCED	IMPCONCN	How much difficulty did you have concentrating?
	IMPGOUT	How much difficulty did you have getting out of the house?
	IMPGOUTM	Did emotional problems keep you from leaving the house?
	IMPPEOP	How much difficulty did you have dealing with people?
	IMPPEOPM	Did emotional problems keep you from dealing with people?
	IMPSOC	How much difficulty did you have participating in social activities?
Alternative Score: WHODASDAED	IMPSOCM	Did emotional problems keep you from participating in social activities?
	IMPHHLD	How much difficult did you have taking care of household responsibilities?
	IMPHHLDM	Did emotional problems keep you from taking care of household responsibilities?
	IMPRES	How much difficulty did you have taking care of responsibilities at work or school?
	IMPRESM	Did emotional problems keep you from working or going to school?
	IMPWORK	How much difficulty did you have getting your daily work done?
	IMPREMEM	How much difficulty did you have remembering to do things?

10.7.2.3.3 Imputation Step

For the categorical outcomes in sets 3 through 9, as well as the worse month outcome in set 1, collocated stochastic imputation was used to replace missing values. There are no deviations from the collocated stochastic imputation procedure described in Section 3.6.3. For more information on the imputation step for these outcomes, see Appendix E.

For the continuous outcomes related to the K6 and WHODAS in imputation sets 1 and 2, a hot-deck step was used to replace missing values. However, although the scale scores were used for modeling, once a donor was selected, the donor's value(s) for each individual question comprising the relevant scale (see [Tables 10.8](#) and [10.9](#)) was used to replace any missing item(s) in the recipient's record for each of the scale components (i.e., multivariate assignment). The final K6 and WHODAS scores were then calculated based on these imputation-revised variables for the individual questions. This process was used to ensure the total scores are consistent with any nonmissing values for respondents who do not have missing data for all items in the scale as well as consistent with each other.

The imputation steps for the K6 (past month and past year) did not include any logical constraints on the donor pool. In addition to the delta constraint, likeness constraints for the K6 imputations required the donor's values to match the recipient's values for any nonmissing items included in the scale. Upper and lower bounds on the total score were also calculated for each nonrespondent, and a likeness constraint required the donor's untransformed total score to be within these bounds. Additional details on the K6 imputation steps are presented in Appendix E.

The imputation step for the WHODAS variables included logical constraints to ensure consistency across variables. For example, if individuals reported that they did not leave the house on their own during the one month when their emotions, nerves, or mental health interfered most with their daily activities (IMPGOUT=5), but they did not answer the follow-up question (IMPGOUTM=missing), then the donor pool was restricted to individuals who also reported that they did not leave the house on their own. This constraint ensures that the recipient receives a value for the follow-up question that is consistent with the gate question. In addition to the delta constraint, likeness constraints were also used that required the donor's values to match the recipient's values for any nonmissing items in the scale. Upper and lower bounds on the total score and the alternative score were also calculated for each nonrespondent, and a likeness constraint required the donor's untransformed score values to be within these bounds. Additional details on the WHODAS imputation step are presented in Appendix E.

10.7.2.3.4 Final Variables

The final imputation-revised variables were identified with the prefix IR followed by the edited variable name. Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were the standard levels used for most imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically

imputed, and 9 = legitimate skip. The imputation-revised variables related to adult mental health and adult depression are shown in [Table 10.8](#).

Table 10.8 Imputation-Revised Variables

Variable Group	Imputation-Revised Variables
Kessler-6	IRDSTCHR30, IRDSTEFF30, IRDSTHOP30, IRDSTNGD30, IRDSTNRV30, IRDSTWORST, IRDSTCHR12, IRDSTEFF12, IRDSTHOP12, IRDSTNGD12, IRDSTNRV12
World Health Organization Disability Assessment Schedule	IRIMPCONCN, IRIMPGOUT, IRIMPGOUTM, IRIMPPEOP, IRIMPPEOPM, IRIMPSOC, IRIMPSOCM, IRIMPHLD, IRIMPHLDM, IRIMPRES, IRIMPRESM, IRIMPWORK, IRIMPREMEM
MDE	IRAMDELT, IRAMDEYR, IRAMDEIMP
Adult Suicidal Thoughts and Behaviors	IRSUICTHNC, IRSUIPLANYR, IRSUITRYR
Adult Suicidal Thoughts and Behaviors due to COVID-19	IRCOSUITHNC, IRCOSUIPLNYR, IRCOSUITRYR
Medical Attention for Suicide Attempt	IRSUIMEDYR
Hospital Stay for Suicide Attempt	IRSUIHOSPYR

COVID-19 = coronavirus disease 2019; MDE = major depressive episode.

10.7.3 Recodes for Additional Analyses

From the final imputation-revised variables, some additional recodes were created to use in additional analyses. [Table 10.9](#) contains information on these additional recoded variables. For information on the creation of these recodes, see the *2023 National Survey on Drug Use and Health Public Use File Codebook* (Center for Behavioral Health Statistics and Quality, 2024b).

Table 10.9 Recoded Variables

Variable Name	Variable Description
KSSLR6MON	K6 TOTAL SCORE IN PAST MONTH
SPDPSTMON	PAST MONTH SERIOUS PSYCHOLOGICAL DISTRESS INDICATOR
KSSLR6YR	K6 TOTAL SCORE IN PAST YEAR
KSSLR6MAX	WORST K6 TOTAL SCORE IN PAST YEAR
SPDPSTYR	PAST YEAR SERIOUS PSYCHOLOGICAL DISTRESS INDICATOR
AKSSLR6WRST	ALTERNATIVE WORST K6 TOTAL SCORE
WHODASTOTSC	WHODAS TOTAL SCORE
WHODASDASC	ALTERNATIVE WHODAS SCORE

11. Editing and Imputation for the NSDUH Roster Variables

11.1 Introduction

This chapter describes the techniques used to edit and impute variables associated with the household roster for the National Survey on Drug Use and Health (NSDUH). The variables described in this chapter can be divided into three groups:

- respondent-level detailed roster variables,
- roster-derived household composition variables, and
- proxy variables.

The respondent-level detailed roster variables included the number of individuals living in the household, as well as the age, gender, and relationship to the respondent for each household member. In the web version of the questionnaire, the household roster module was self-administered. In the computer-assisted interviewing (CAI) version, the introductory question for the household roster portion of the questionnaire (QD54) was interviewer administered. If only one person lived in the household or the respondent did not know or refused to answer, then the household composition (roster) section was skipped. Otherwise, the respondent was asked questions about the age, gender, and relationship to the respondent of every member of the household, starting with the respondent and including up to 25 household members.

If two household members were selected to complete the interview, the variable MBRSELECT indicated which member of each respondent's household roster corresponded to the other selected respondent. In the CAI version of the survey, this question was answered by the interviewer rather than the respondent. The roster entry for the respondent was referred to as the "self" entry. In effect, the respondent completed a grid with the number of rows corresponding to the value entered in QD54. [Table 11.1](#) shows an example grid where the number of individuals in the household is four. In this example, the roster of the respondent is shown, and the indicator variable shows that the respondent's son was selected as the other respondent.

Table 11.1 Roster Grid Example Where Number of Individuals in Household (QD54) Equals 4

Individual #	Relationship to Respondent	Age in Years	Other Member Selected for Pair ¹
1	Self	44	0 (No [Impossible])
2	Husband	42	0 (No)
3	Son	16	1 (Yes)
4	Tenant, boarder, or exchange student	16	0 (No)

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

The possible roster relationship codes reported by a respondent are listed in [Table 11.2](#). Because the roster relationship codes were expanded in 2016, [Table 11.2](#) includes the equivalent relationship codes from survey years 2002 through 2015. The levels available from 2002 through 2015 are still used to create comparable variables, specifically ROSRLT1 through ROSRLT25, which are discussed in Section 11.2.

Table 11.2 Detailed Roster Relationship Codes (2016-Present) and Equivalent Former Roster Relationship Codes (2002-2015)

ROSRLTD1- ROSRLTD25 Value	Detailed Relationship to Respondent (2016-Present)	ROSRLT1- ROSRLT25 Value	Relationship to Respondent (2002-2015)
1	Self	1	Self
2	Spouse	5	Spouse
3	Child ¹	3	Child
4	Child-in-law	8	Child-in-law
5	Sibling ¹	4	Sibling
6	Sibling-in-law	13	Other relative ²
7	Parent ¹	2	Parent
8	Parent-in-law	10	Parent-in-law
9	Aunt or uncle	13	Other relative ²
10	Niece or nephew	13	Other relative ²
11	Grandparent	11	Grandparent
12	Grandchild	9	Grandchild
13	Cousin	13	Other relative ²
14	Ex-spouse	13	Other relative ²
15	Unmarried partner	6	Unmarried partner
16	Housemate or roommate	7	Housemate or roommate
17	Tenant, boarder, or exchange student	12	Boarder or roomer
18	Other relative	13	Other relative ²
19	Other nonrelative	14	Other nonrelative

¹ Respondents are instructed to classify step, foster, or adoptive child relationships as “child”; step, foster, or adoptive parent relationships as “parent”; and half, step, foster, or adoptive sibling relationships as “sibling.”

² Several detailed relationship codes were collapsed into the “Other relative” level for the back-coded relationship to respondent measure.

The second group of variables, the roster-derived household composition variables, was derived from the respondent-level detailed roster variables. These were mostly count variables reporting the number of individuals in the roster with various characteristics. These variables and their imputation-revised counterparts are listed in [Table 11.3](#). The family relationships for these variables are determined using the detailed relationship codes introduced in 2016. Among the three groups of variables described in this chapter, only six of the roster-derived household composition variables underwent imputation.

Table 11.3 Roster-Derived Household Composition Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Total number of rostered individuals	TOTPEOP	IRHHSIZE
Number of individuals in household aged 17 or younger	KID17	IRKID17
Number of individuals in household aged 65 or older	HH65	IRHH65
Indicator of whether the respondent had family members in household	FAMSKIP	IRFAMSKIP
Number of respondent's family members in household (includes foster relationships)	FMLYSIZE	IRFMLYSZ
Number of respondent's family members in household aged 17 or younger (includes foster relationships)	KIDFMLY	IRKDFMLY
Number of respondent's children in household aged 2 or younger	NRBABIES	N/A
Number of respondent's children in household aged 3 to 5 years old	NRPRESCH	N/A
Number of respondent's children in household aged 6 to 11 years old	NRYUNGCH	N/A
Number of respondent's children in household aged 12 to 17 years old	NRTEENS	N/A
Number of respondent's children in household aged 17 or younger	NRCH0_17	N/A
Number of respondent's children in household aged 18 to 20 years old	NROLDRCH	N/A
Number of respondent's children in household aged 21 or older	NROLDCH	N/A
Number of roommates/housemates in household	NROOMATE	N/A
Indicator of presence of mother in household (12- to 17-year-olds) ¹	IMOTHER	N/A
Indicator of presence of father in household (12- to 17-year-olds) ¹	IFATHER	N/A

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

The third group of variables, the proxy variables, allowed for the selection and identification of another household member to answer the health insurance coverage and income questions for the respondent. This proxy respondent was required to be a relative of the respondent and aged 18 years or older. The proxy module was also available on the web version of the questionnaire. The edited versions of these variables and the questions to which they map are shown in [Table 11.4](#).

Table 11.4 Proxy Variables

Raw Variable	Text of Survey Question Associated with Raw Variable	Edited Variable
QP01	Is there anyone else who lives here who is 18 or older who would be better able to give me the correct information about your health insurance coverage and the kinds of income you receive?	PRXABLE2
QP02	Who is the person you think can help us get the correct information for these questions?	PRXRELATD PRXRELAT
QP03	Is your [QP02 fill] available right now?	PRXHOME2
QP04	Would you ask your [QP02 fill] to join us to help with these last questions about health insurance and income?	PRXJOIN2
HASJOIN	Has the person's [QP02 fill] joined R?	PRXYANS2
QP03a	Is there any other adult family member available who might be able to answer these questions?	PRXRETRY

11.2 Editing the Respondent-Level Detailed Roster Variables

This section describes the methods used to create edited versions of the respondent-level detailed roster variables: ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), ROSRLTD1-ROSRLTD25 (detailed relationship to respondent), ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only), and TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither). These variables describe up to 25 members of the household.

The editing procedures for the respondent-level detailed roster variables began with consistency checks included in the Blaise program code, which were implemented to reduce the amount of editing required at the data processing stage. The consistency checks in the questionnaire were supplemented with other edits involving the respondent-level detailed roster variables outside the CAI instrument. These involved resolving instances where it was unclear which roster member was the self and also where relationship codes were impossible (or very unlikely) given the age and gender in relation to the self. The web version of the NSDUH questionnaire omits the consistency checks from the household roster module to reduce respondent burden. The household roster editing programs are set up to duplicate the logic of the consistency checks that exist in the in-person CAI version of the questionnaire, so the editing processes for relationship codes, gender, and ages of roster members remained the same.

Section 11.2.1 describes the consistency checks programmed into the CAI version of the questionnaire. Section 11.2.2 describes the creation of a roster-level dataset for further processing. Section 11.2.3 describes roster edits involving the self. Section 11.2.4 describes roster edits for other household members, after the self has been established. Finally, Section 11.2.5 describes the creation of the final edited respondent-level detailed roster variables.

11.2.1 Roster Consistency Checks

Two types of consistency checks were employed in the CAI instrument for the household roster section of the questionnaire. These checks (1) compared the roster entry corresponding to the respondent with previously entered questionnaire information and (2) compared a roster entry against other roster entries for internal consistency. The interviewer could override the consistency checks and explain why the response given was correct. These consistency checks and override responses were not included in the web instrument.

11.2.1.1 Comparisons with Previously Entered Questionnaire Information

Gender and age were the two consistency checks built into the household roster section of the CAI instrument that compared the roster entry for “the self” with the previously entered questionnaire information. The check for gender was triggered if the respondent entered a gender in the household roster that was different from the one previously recorded in the interview (question QD01). If the gender did not match, the interviewer was required to change either the roster entry or QD01.

The consistency check for age compares the respondent’s age in the roster with the age previously entered in the questionnaire (the Blaise variable CURNTAGE). The interviewer could either change the respondent’s age entered in the roster or override the consistency check. If the interviewer chose to override the consistency check, then the interviewer provided an explanation as to why the roster age did not match CURNTAGE. Explanations given by the interviewer for overriding this consistency check were carefully reviewed. Occasionally, the final value for age (AGE) was set to the age of the self in the household roster (the “roster age”) based on these explanations as well as other evidence. Additional details about how roster age was used for creating AGE are described in Section 4.2.2. Strategies for the more common situation, where the original value for AGE was not set to the roster age, are discussed in Section 11.2.3.

11.2.1.2 Internal Consistency Checks

Internal consistency checks were implemented in the CAI instrument to compare one roster entry with another. These checks are also mirrored by the roster editing SAS® programs, which ensure web data are edited in a manner consistent with the CAI data. These checks were triggered if any of the following conditions occurred in the CAI instrument:

1. The interviewer or respondent reported that the respondent had more than one spouse or unmarried partner or reported a spouse and an unmarried partner.
2. The interviewer or respondent reported having a grandparent who was the same age or younger than the respondent.
3. The interviewer or respondent reported having a grandchild who was the same age or older than the respondent.
4. The interviewer or respondent reported a spouse, an unmarried partner, or an ex-spouse who was 16 years old or younger.

5. The interviewer or respondent reported that the respondent had a spouse, unmarried partner, or ex-spouse and the respondent was 16 years old or younger.
6. The interviewer or respondent reported that the respondent was either a child-in-law or a parent-in-law and the respondent was 16 years old or younger.
7. The interviewer or respondent reported having a child-in-law who was the same age or older than the respondent.
8. The interviewer or respondent reported having a parent-in-law who was the same age or younger than the respondent.
9. The interviewer or respondent reported having a grandparent or grandchild and the age difference between the respondent and the household member was less than 30 years.
10. The interviewer or respondent reported that only one member of a pair was available for interviewing, potentially due to a change in household size.

Usually, if a consistency check was triggered, the interviewer changed either an age or a relationship code in the roster to a more appropriate value. In some situations, the interviewer overrode the consistency check. Overridden consistency checks were reviewed during the data processing stage. The relevant household roster and the explanation given by the interviewer for the override were examined to determine whether the override was legitimate. If the override was deemed legitimate (e.g., a father marries a woman, listed as mother, who is younger than the respondent), the original answer was allowed to remain and no edit was applied. If the interviewer's explanation was not considered legitimate, then an edit was applied. More details about roster edits are provided in Section 11.2.4.

11.2.2 Roster-Level Dataset

To facilitate processing of the roster variables, a roster-level dataset was created in which the number of records per respondent was given by the household size in question QD54. For example, if a respondent indicated a household of three consisting of the respondent, a mother, and a father, then there would be three records on the dataset associated with this respondent: one for the self, one for the mother, and one for the father. Even if the respondent did not start or complete the roster questions, records were created for each household member. If the roster was completely missing, including the number of household members, a household roster consisting of one person, the respondent, was created. If a pair of respondents from one household were interviewed and the roster only listed one person, a second record was added to the roster-level file to hold the data from the other pair member.

11.2.3 Roster Edits Involving the Self

The Blaise program code used by the CAI instrument required the interviewer to identify exactly one self and a corresponding age and gender in the household roster. Moreover, the interviewer was required to confirm with the respondent that the respondent was in fact the identified self. Because the check involving gender was not allowed to be overridden, the gender for the self in the roster always matched QD01, which was equivalent to IRSEX (Section 4.2.4). However, it was possible to have problems matching AGE with the age of the self in the roster, despite the consistency check comparing the respondent's roster age against CURNTAGE (see Section 4.2.2 for a description of the methodology used to create AGE).

The interviewer was able to override the consistency check for age of the self for one of two reasons: (1) the self was misidentified and another roster member was the true self but the interviewer decided not to change the entries; or (2) the interviewer correctly identified the self but indicated that the correct age for the respondent was different than CURNTAGE, and other evidence did not support this claim (AGE was not set to the roster age, as discussed in Section 11.2.1.1).

11.2.3.1 Original Self Misidentified: Identifying the Real Self

If the self was misidentified in the roster, an attempt was made to identify a self among the other roster members. A roster member was selected as the self under one of two possible circumstances: (1) the roster member's age, gender, and relationship data were missing; or (2) the roster member was of the respondent's gender and was within 1 year of the respondent in age.¹⁷¹ If more than one roster member met the above criteria, one of these roster members was chosen to be the self because all subsequent editing steps required that there be exactly one self on the roster.

In some cases, the respondent did not report a self at all. There are three editing procedures implemented to identify a self that corresponded with the respondent in the roster:

1. If the self was missing from the roster, and the other pair member or screener data listed more household members than the respondent reported in QD54, a record for the self was added to the roster, and the age and gender were set to IRSEX and IRAGE.
2. If the self was missing from the roster, and the roster had a household member with partial or complete missing values, the missing row was assigned as the self, and the missing age and/or gender data were set to IRSEX/IRAGE.
3. If the self was missing from the roster, and the roster had exactly one set of duplicate household members reported, and the respondent could be identified on the screener, the duplicate was edited to be the self, and the age and/or gender were set to IRSEX/IRAGE.

After identifying the self, the remaining relationship codes were set to bad data codes because it was assumed that the interviewer used the roster member identified as the self, rather than the respondent, as the point of reference.

11.2.3.2 Original Self Misidentified: Salvaging Relationship Codes

If the self was misidentified, all other relationship codes were set to missing because the reference person was someone other than the respondent. Most of the time, however, the original relationship codes were salvaged by reviewing the suppression messages from the interviewer and by comparing the respondent's roster to the screener roster and to the other pair member's roster, when available. If these other data sources were not informative, relationship codes could still be salvaged under the following circumstances:

¹⁷¹ A 1-year difference was allowed because the respondent's age might have changed during the interview. In this instance, the values of AGE and CURNTAGE may have differed by 1 year.

1. If the reference person was the respondent's sibling, the roster member listed as "self" was actually a sibling, and most other relationship codes were salvaged. (Generally, relationships between the respondent and other household members would be the same with a sibling. For example, the respondent's parents are also the respondent's sibling's parents.)
2. If the reference person was the respondent's spouse or unmarried partner, the roster member listed as "self" was changed to spouse or unmarried partner, and the children relationship codes were salvaged.
3. If all the roster members other than the misidentified self were either roommates, boarders, or other nonrelatives, then the reference person was the respondent's roommate, boarder, or other nonrelative. All other relationship codes were salvaged.

11.2.4 Roster Edits for Other Household Members

Before editing the roster entries for other household members, an attempt was made to fill in any missing or bad data in the roster using either the screener roster or the other member of a responding pair, when available. The difficulty with comparing rosters between the respondent and the screener and between the respondent and the other pair member is that they do not all have the same person of reference. For example, consider a household with a 55-year-old male, his 52-year-old wife, and their two children, a 16-year-old boy and a 12-year-old girl. [Table 11.5](#) shows the relationship codes for each person of reference. The columns for the relationship codes cannot be compared directly. However, the collection of columns, or roster matrix, is fixed for the entire household. Because relationships are symmetrical (e.g., parent/child, spouse/spouse, sibling/sibling), most of the matrix is frequently defined by just one column.

Table 11.5 Relationship Codes for Each Possible Person of Reference

Person	Age	Gender	Person of Reference			
			1	2	3	4
1	55	Male	1 – Self	2 – Spouse	7 – Parent	7 – Parent
2	52	Female	2 – Spouse	1 – Self	7 – Parent	7 – Parent
3	16	Male	3 – Child	3 – Child	1 – Self	5 – Sibling
4	12	Female	3 – Child	3 – Child	5 – Sibling	1 – Self

Editing these matrices consists of two parts. First, the household matrix, like that shown in [Table 11.5](#), is created from the reported household roster, with the roster members sorted by age from oldest to youngest, and columns are created to show all possible relationships between all household members. Matrices are created for each respondent and for the screener. Second, the program compares the matrix for a respondent with the screener matrix and with the other pair member's matrix and fills in missing values when possible. For example, in the roster shown in [Table 11.5](#), if the relationship code, age, or gender for the respondent's sibling was missing, but the respondent's sibling was present on either the other pair member data or screener data and had a complete record, then the missing roster record for the

respondent's sibling would be updated using the other pair member or screener data. When the other pair member data are present, preference is given to data from the other pair member over the screener data. The screener roster can only be used to fill in missing data for household members aged 12 or older. Nonmissing ages, genders, or relationship codes are not changed at this step, nor is household size changed to be consistent with the other pair member or the screener. This editing is performed prior to any editing of the household members other than the self so that the household roster is as complete as possible before further edit rules are applied.

After filling in missing data where possible, the entire roster was checked for internal consistency. Relationship codes were edited if the relationship of the roster member to the self was logically impossible based on age and gender. Edits of household roster ages, genders, and/or relationship codes either changed the reported value to another valid value or changed the reported value to bad data. For example, a respondent may list multiple children-in-law but no children. Frequently, these are stepchildren or a partner's children and should be classified as either children or other nonrelatives. These edits are detailed in Appendix G.

The majority of these edits were manually reviewed for accuracy. Some edits were associated with consistency checks, and interviewers' explanations for overrides to these consistency checks were carefully examined to assess the legitimacy of the override as explained in Section 11.2.1. Some edits were "automatic" in the programming code, in that the interviewer was assumed to have been incorrect when the override was implemented. These edits were undone if the interviewer's explanation for the override was considered legitimate. In other situations, the default strategy was to assume that the override of the consistency check was correct, and, therefore, the edit was applied only if the interviewer's explanation appeared incorrect. For example, if a respondent lists three or more parents, those relationship codes are automatically set to bad data. However, there are households where a child may live with biological parents and a stepparent. If the interviewer indicated that this was the correct situation, or if the other pair member's roster agreed, the codes were changed back to their original values.

There were two additional edits that were not reviewed by hand. These were edits involving the other pair member.

1. If two members were selected in a household, the roster age for the other member selected was commonly not the same as the questionnaire-edited age (AGE, defined in Section 4.2.2) of the other pair member. In this situation, the roster age for the other member selected was changed to this questionnaire-edited age value.
2. If two members were selected in a household, the gender that one member selected for the other on the household roster was often not the same as the gender (IRSEX, defined in Section 4.2.4) reported by that other pair member. In this situation, the roster gender was changed to match the gender value that the other pair member reported.

These edits were not reviewed because they happen quite frequently, and because the final AGE and IRSEX values from the other pair member were used, they were assumed to be accurate. [Table 11.6](#) lists the number of times these changes occurred in 2023.

Table 11.6 Changes to Roster Age or Gender Based on the Other Pair Member Reported

Variable Changed	Description	Count
Age	Difference in other pair member's roster age and AGE is 2 years or less	2,411
	Difference in other pair member's roster age is greater than 2 years	238
Gender	Other pair member's roster gender does not match IRSEX	207

11.2.5 Final Edited Respondent-Level Detailed Roster Variables

The raw roster variables contained the following information for each roster member: age, gender, relationship to respondent, and a 0/1 variable that indicated whether the roster member was the other member selected in a pair (see [Table 11.1](#) for an example). This information could be captured for up to 25 members of a household. Within the CAI and web instruments, separate variables were created to collect this information for male and female household members and for household members with ages reported in years as opposed to months. When the edited versions of these variables were created, this information was combined for each household member into four variables, one for each attribute (i.e., age, gender, relationship to respondent, and pair status). After completing the edits listed in Section 11.2, the roster-level file was transposed back to a respondent-level file. The edited age, gender, relationship, and pair status variables were used to create the detailed roster variables: ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), ROSRLTD1-ROSRLTD25 (detailed relationship to respondent), and ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only). Additional variables created were TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither). The twin-type variables came directly from the questionnaire and were not edited. Final edited versions of the respondent-level detailed roster variables were used to derive (or, at a minimum, to calculate bounds when data were missing) the household composition variables described in Section 11.3.

11.3 Editing and Imputation Procedures for Roster-Derived Household Composition Variables

11.3.1 Creation of Edited Roster-Derived Household Composition Variables

This section discusses the creation of edited versions of the roster-derived household composition variables shown in [Table 11.3](#). These counts were created after editing of the household roster was complete. If any information in the roster was missing, the roster-derived variable was set to missing. However, if only some of the roster records for a respondent's household had missing data, then roster records with nonmissing data for that household were used to create bounds for imputation. Details on the imputation of the roster-derived household variables are provided in Section 11.3.2. Note that editing and imputation of these variables is done at the respondent level rather than at the household level. That is, if two respondents

were selected in a single household as part of a pair, then no effort was made to reconcile the counts across the two respondents.

The respondent's household size was equal to the total number of rostered individuals in the household, TOTPEOP, as shown in [Table 11.3](#). The value of TOTPEOP was expected to equal the value of QD54 in most situations. However, sometimes the original self was misidentified and no other roster members were close to matching the respondent's age and gender. In these circumstances, an extra roster member was added to correspond to the respondent (the self) so that the value of TOTPEOP was 1 greater than the value of QD54. For other situations, the respondent did not enter a value for QD54, and thus TOTPEOP and all the roster-derived variables were missing. Finally, it was possible that duplicate entries were put into the household roster so that the value of TOTPEOP could be determined by excluding the duplicates from the roster. This latter situation was usually impossible to detect, unless the respondent had two fathers, two mothers, or two spouses of exactly the same age. In this instance, the respondent's roster was compared with the roster of the other pair member (OPM) and the screener roster, when available.

The variables KID17 (number of individuals in the household aged 17 or younger) and HH65 (number of individuals in the household aged 65 or older) were simple counts based on the roster ages and did not account for the relationships of the individuals to the respondent. If some of the roster members had missing ages, then the values of KID17 and HH65 also were missing, regardless of whether any of the other roster members were eligible to be part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in Section 11.3.2.2. However, if the roster member was missing a relationship code, but not an age, then that roster member was still eligible to be included in these counts.

The variable FAMSKIP was an indicator of whether the respondent's household contained any other family members. It was created based on the relationship codes of the roster members. If one or more of the roster members had a missing relationship code, and no other family members were in the respondent's household, then the value of FAMSKIP was set to missing. However, if one of the nonmissing roster members' relationship codes indicated that the household contained one of the respondent's family members, then the value of FAMSKIP was not missing, even if other roster members had missing relationship codes.

The variables FMLYSIZE (number of respondent's family members in the household, including foster relationships) and KIDFMLY (number of respondent's family members in the household aged 17 or younger, including foster relationships) were simple counts based on the relationships of the individuals to the respondent and the ages in the respondent's household roster. FMLYSIZE and KIDFMLY were created to determine appropriate measures of poverty levels, using Federal poverty definitions. If some of the roster members had missing ages or missing relationship codes, the values of FMLYSIZE and KIDFMLY were set to missing, even though some of the roster members might have been eligible to be part of the count. In these instances, the imputed values were restricted based on the nonmissing information available in the roster, as explained in Section 11.3.2.2.

Ten other roster-derived variables were created that used both the age and relationship codes of the roster members. All of the roster-derived variables and their definitions are summarized in [Table 11.3](#). Except for FAMSKIP, each of these variables was missing if the age or relationship codes for at least one roster member in a respondent's household were missing. FAMSKIP could be coded despite missing values if there was at least one nonmissing family relationship code in the roster. Edited versions of the roster-derived variables were also used in the editing procedures applied to the creation of the edited proxy variables, described in Section 11.4.

11.3.2 Imputation for Roster-Derived Household Composition Variables

Of the three groups of edited roster variables described in the introduction to this chapter, the only group that underwent any imputation at all was the following subset of the roster-derived household composition variables: TOTPEOP, KID17, HH65, FAMSKIP, FMLYSIZE, and KIDFMLY. Each of these six variables formed its own single-member imputation set¹⁷² and tended to have item response rates of more than 95 percent. The single RP/single PRD type of PMN, described in Section 3.4.1, was used to impute nonmissing values among these six variables in the order shown in [Table 11.3](#). The order was important, as imputation-revised variables from earlier in the sequence were frequently used to assist with imputation-revised variables later in the sequence.

Section 11.3.2.1 describes the imputation process applied to the first edited variable, TOTPEOP. Section 11.3.2.2 summarizes the imputation processes applied to the other five variables. Because the processes applied to the other five variables are very similar to the process applied to TOTPEOP, Section 11.3.2.2 only lists divergences from the process that was applied to the TOTPEOP variable.

11.3.2.1 Imputation for TOTPEOP (Imputation Set 1)

The first imputation set included a single variable, TOTPEOP. The corresponding imputation-revised variable IRHHSIZE was created using the single RP/single PRD type. There were no noteworthy deviations from this general approach. Section 11.3.2.1.1 describes the RP step, Section 11.3.2.1.2 describes the PRD step, and Section 11.3.2.1.3 describes the hot-deck step.

11.3.2.1.1 Response Propensity Step

The response propensity model for imputation set 1 used the preliminary analysis weight, PANALWT. All respondents were in the domain for the TOTPEOP variable. A domain member was considered an item respondent if and only if TOTPEOP was nonmissing. See Appendix D for details on the covariates used in the RP models for this variable.

¹⁷² An imputation set is a set of variables for which a single donor is used in the final hot-deck step. Chapter 3 describes this concept more fully.

11.3.2.1.2 Prediction Step

TOTPEOP is a count variable. It was assumed to have a Poisson distribution, and the parameters for the models were estimated using the adjusted weights that are outputs of the RP step and using Poisson regression as implemented by the LOGLINK procedure in SUDAAN software.¹⁷³ The single predicted mean used in the subsequent hot-deck step was the predicted number of people in the household.

11.3.2.1.3 Hot-Deck Step

The hot-deck step for the TOTPEOP variable was the simplest one used in the NSDUH. There were no logical constraints, and the only likeness constraint was the delta constraint. The predictive mean vector was a scalar. Every item nonrespondent was handled on the first attempt to find a donor. Additional details on the hot-deck step for TOTPEOP are available in Appendix E.

11.3.2.2 Imputation for Other Roster-Derived Household Composition Variables That Underwent Imputation (Imputation Sets 2 through 6)

Like TOTPEOP, the remaining five roster-derived household composition variables that underwent imputation (from [Table 11.3](#): KID17, HH65, FAMSKIP, FMLYSIZE, and KIDFMLY) were handled separately using the single model type of PMN and used the preliminary analysis weight, PANALWT, in the RP step. The methods were very similar, with only a few exceptions, as follows:

- FAMSKIP was a dichotomous variable, not a count variable. Therefore, its PRD model was a logistic regression model as implemented by the RLOGIST procedure in SUDAAN. The single predicted mean used in the later hot-deck step was the predicted probability that the respondent did not live with any family members.
- Bounds were determined for every other variable except FAMSKIP. These bounds were based both on nonmissing information in the roster and on previously imputed variables. For each of these variables, a single logical constraint was used in the hot-deck step, which required the donor to have a value within the bounds.
- Previously imputed roster-derived household composition variables were frequently used in likeness constraints in the hot-deck steps.

Appendix D provides details on the covariates used in the RP models for these variables.

11.4 Editing the Proxy Variables

This section describes the creation of edited proxy variables, as listed in [Table 11.4](#). Section 11.4.1 describes the creation of an indicator variable, EDFAM18, which was used to determine

¹⁷³ Details about the LOGLINK procedure and additional references are provided in the *SUDAAN® Language Manual, Release 11.0.1* (RTI International, 2013).

skip patterns and missing codes for the five edited proxy variables. Sections 11.4.2 and 11.4.3 describe the editing processes for each value of EDFAM18.

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

1. There was more than one person in the household.
2. The eligible person was a relative (not a boarder, roommate, or some other nonrelative).
3. The eligible person was aged 18 or older.

The correspondence between the six questionnaire items in the proxy section of the questionnaire and the corresponding edited variables are shown in [Table 11.4](#). Except for QP02 and its edited variables PRXRELAT and PRXRELATD, the valid questionnaire responses were “1 = Yes” and “2 = No.” QP02 and PRXRELATD had multiple responses ranging from 1 to 31, with each level representing the relationship of the proxy to the respondent. PRXRELATD was introduced in 2016 to capture the detailed relationship codes available as responses to QP02. PRXRELAT is still created for consistency between survey years by recoding the 31 values available from PRXRELATD to the 21 levels available prior to 2016. The levels of PRXRELAT are shown in [Table 11.7](#), and the levels of PRXRELATD are shown in [Table 11.8](#). Depending on household size and composition, QP02 gave respondents a maximum of nine possible proxy choices out of the 31 relationship types detailed by PRXRELATD. In addition to the nine possible answer choices presented by QP02 based on reported household size and composition, respondents were also given a 10th and final choice of “other adult relative” to select a proxy who is related to the respondent but may not live in the household. Among the detailed relationship codes introduced in 2016, roster members reported as sibling-in-law, uncle/aunt, nephew/niece, cousin, or ex-spouse were considered relatives and thus were eligible for selection as a proxy.

Table 11.7 Levels of PRXRELAT

PRXRELAT	Relationship of Proxy Member	Gender of Proxy Member
1 = Father	Parent	Male
2 = Mother	Parent	Female
3 = Son	Child	Male
4 = Daughter	Child	Female
5 = Brother	Sibling	Male
6 = Sister	Sibling	Female
7 = Husband	Spouse	Male
8 = Wife	Spouse	Female
9 = Male Unmarried Partner	Unmarried partner	Male
10 = Female Unmarried Partner	Unmarried partner	Female
11 = Son-in-law	Child-in-law	Male

Table 11.7 Levels of PRXRELAT (continued)

PRXRELAT	Relationship of Proxy Member	Gender of Proxy Member
12 = Daughter-in-law	Child-in-law	Female
13 = Grandson	Grandchild	Male
14 = Granddaughter	Grandchild	Female
15 = Father-in-law	Parent-in-law	Male
16 = Mother-in-law	Parent-in-law	Female
17 = Grandfather	Grandparent	Male
18 = Grandmother	Grandparent	Female
19 = Other Male Relative	Other relative	Male
20 = Other Female Relative	Other relative	Female
21 = Other Adult Relative	Other relative	Male or Female

Table 11.8 Levels of PRXRELATD

PRXRELATD	Relationship of Proxy Member	Gender of Proxy Member
1 = Husband	Parent	Male
2 = Wife	Parent	Female
3 = Son	Child	Male
4 = Daughter	Child	Female
5 = Son-in-law	Child-in-law	Male
6 = Daughter-in-law	Child-in-law	Female
7 = Brother	Sibling	Male
8 = Sister	Sibling	Female
9 = Brother-in-law	Sibling-in-law	Male
10 = Sister-in-law	Sibling-in-law	Female
11 = Father	Parent	Male
12 = Mother	Parent	Female
13 = Father-in-law	Parent-in-law	Male
14 = Mother-in-law	Parent-in-law	Female
15 = Uncle	Uncle/aunt	Male
16 = Aunt	Uncle/aunt	Female
17 = Nephew	Nephew/niece	Male
18 = Niece	Nephew/niece	Female
19 = Grandfather	Grandparent	Male
20 = Grandmother	Grandparent	Female
21 = Grandson	Grandchild	Male

Table 11.8 Levels of PRXRELATD (continued)

PRXRELATD	Relationship of Proxy Member	Gender of Proxy Member
22 = Granddaughter	Grandchild	Female
23 = Male cousin	Cousin	Male
24 = Female cousin	Cousin	Female
25 = Ex-husband	Ex-spouse	Male
26 = Ex-wife	Ex-spouse	Female
27 = Male unmarried partner	Unmarried partner	Male
28 = Female unmarried partner	Unmarried partner	Female
29 = Other Male Relative	Other relative	Male
30 = Other Female Relative	Other relative	Female
31 = Other Adult Relative	Other relative	Male or Female

11.4.1 Edited Indicator of Potential Proxies in Household (EDFAM18)

As described in Section 11.3.1, a binary variable (FAMSKIP) was created that indicated whether the respondent's household roster included other family members. If the presence or absence of other family members was unknown because of a missing household size or missing values in the roster, FAMSKIP could not be determined. A similar variable was created to identify households where the respondent's household roster included other family members aged 18 years or older ("adult" family members), any one of whom could potentially serve as a proxy for the respondent. The edited indicator was called EDFAM18, where "1" indicated no potential proxy existed in the respondent's household, "0" indicated a potential proxy did exist in the respondent's household, and "98" indicated unknown.

11.4.2 Editing the Proxy Variables when EDFAM18 = 1

In general, a value of EDFAM18 = 1 implied that the respondent was skipped out of the proxy questions because no potential proxy existed in the household. In these circumstances, all of the proxy variables were given a legitimate skip code (99). Two situations could occur, however, where adult family members were incorrectly identified in the household roster: (1) the respondent had not identified any adult family members in the household but had nonfamily members in the household whose ages were not known; and (2) the raw household roster indicated that one potential proxy existed in the household but editing changed the age of this single potential proxy to younger than 18. In these circumstances, the respondent was allowed to answer the proxy questions even though the value of EDFAM18 was 1 (i.e., the final edited household roster indicated that no potential proxy existed in the respondent's household). Moreover, in these situations, the interviewer indicated that none of these household members who were incorrectly identified as adult family members were proxies. However, the "no" value in the first raw proxy variable (QP01) was replaced and assigned a code of 89 (LEGITIMATE SKIP, Logically assigned) in the corresponding edited variable (PRXABLE2). When PRXABLE2 was set to 89 (LEGITIMATE SKIP, Logically assigned), all of the

edited proxy variables corresponding to the raw proxy variables, which followed QP01, were assigned a code of 99 (LEGITIMATE SKIP). These were instances in which the respondent answered the proxy section (questions about a proxy) but the interviewer indicated that they were not proxies, so no proxy should have actually answered the health insurance and income sections.

11.4.3 Editing the Proxy Variables when EDFAM18 = 0

If EDFAM18 was 0, the proxy variables were edited as follows:

1. If the raw proxy variables had legitimate nonmissing values (i.e., not replaced by a logically assigned legitimate skip), the edited proxy variables (except PRXRELAT and QP03A) were set to those nonmissing values.
2. If any of the raw proxy variables (except PRXRELAT and QP03A) had a value of 2 ("no"), then all of the variables that followed were edited to legitimate skips.
3. If any of the raw proxy variables (except QP03A) had a value of "don't know" or "refused," then the corresponding edited variable and all the edited variables that followed were given a code of 94 (don't know) or 97 (refused).
4. If any of the raw proxy variables were missing and a legitimate skip code could not be applied, then the corresponding edited variable and all the variables that followed were given a code of 98 (blank).

11.4.4 Additional Editing for PRXRELAT

More detailed rules were used to assign values to PRXRELAT. The value of QP02, which identified the proxy for the respondent, was chosen directly from the respondent's household roster with one caveat. A list of adult family members (a proxy roster) was shown to the respondent, and the respondent was asked to select the family member who could best answer the health insurance and income sections. When the proxy roster included a large number, only the first nine adult family members were listed. Regardless of roster size and composition, respondents were given the final choice to select an "other adult relative" as a proxy, even if that person was not part of the household roster. Once the proxy roster was established, the number selected in QP02 was matched to the corresponding person in the proxy roster.

11.4.5 Additional Editing for QP03A

Item QP03A allowed respondents to select another proxy from the list of potential proxies in QP02 if the previously selected proxy was not available to join the interview. Respondents were redirected to question QP03A if they answered "no" to any of the raw proxy variables; otherwise, QP03A was edited to a skip code. The edited variable of QP03A is PRXRETRY.

12. Editing and Imputation for the NSDUH Income Variables

12.1 Introduction

This chapter describes procedures for editing and imputing the income variables in the 2023 National Survey on Drug Use and Health (NSDUH) interview. The income section is interviewer administered for in-person respondents and self-administered for web respondents. At the end of the income section, questions also are included about cell phone coverage in households. These cell phone coverage variables are edited but not imputed.

Unlike other sections of the interview that asked about behaviors or situations in the 12-month period prior to the interview, the reference period for the income section in the 2023 NSDUH was the previous calendar year (i.e., January 1, 2022, through December 31, 2022). Respondents (or other household members serving as proxies¹⁷⁴) were asked whether they (or other family members living in the dwelling unit, if applicable) received income or benefits from specific sources, such as social security during the previous calendar year. These questions are subsequently referred to as “source-of-income” questions in this chapter. In addition, the income section of the interview included questions about the dollar amounts of total personal income and total family income (if applicable) during the previous calendar year.

Respondents were asked about personal and family-level income from specific sources as well as binary (i.e., \$20,000 or more or less than \$20,000) and finer categories of total annual income at both the personal and family levels. Respondents with no other family members living in their dwelling unit answer all questions about their personal income. A comparison of the personal level and family level of income questions is shown in [Table 12.1](#).

Table 12.1 Comparison between Personal Level and Family Level of Income Questions

Income Questions Included in NSDUH	Personal Level	Family Level
Social Security	No	Yes
Supplemental Security Income	No	Yes
Welfare Payments	No	Yes
Other Welfare Services	No	Yes
Food Stamps/SNAP Benefits	No	Yes
Months on Welfare	No	Yes
Binary Total Income	Yes	Yes
Finer Category Total Income	Yes	Yes

¹⁷⁴ Reference is made only to “respondents” in the remainder of this chapter. However, readers are advised that the income information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the income questions for the respondent. See Chapter 11 for more information about proxy variables.

Consistent with most of the imputation-revised variables that were discussed in the previous chapters of this report, imputations for the income variables were performed using the predictive mean neighborhood (PMN) methodology detailed in Chapter 3. The edits that were applied to the income variables prior to imputation are described in Section 12.2.

12.2 Editing for Binary and Finer Category Income Variables

12.2.1 Binary Income Variables

The source-of-income variables recorded binary (i.e., yes/no) responses. Information for most of these variables was captured through single questions about a given source of income. One exception to this single-question format was for measurement of the number of months on welfare (questions QI12AN and QI12BN and the edited variable WELMOS). Question QI12AN was asked if respondents did not report any family members' receipt of Supplemental Nutrition Assistance Program (SNAP) benefits in question QI07N; otherwise, question QI12BN was asked. In asking about the number of months on welfare, question QI12BN included the additional phrase, "not including SNAP benefits."

Creation of WELMOS also was dependent on respondents' answers in the questions corresponding to the relevant edited variables for family members' receipt of cash assistance from a state or county welfare program (FAMPMT) and receipt of other (noncash) welfare or public assistance (FAMSVC). If FAMPMT or FAMSVC indicated that respondents received welfare payments, WELMOS was assigned the values from questions QI12AN and QI12BN. Otherwise, if family members in the household (including the respondent) definitely did not receive welfare payments or noncash benefits (e.g., job training, help with child care), WELMOS was assigned a legitimate skip code.

In addition to the source-of-income variables, binary variables were created from a pair of questions that asked whether the respondent's personal total income or the respondent's total family income was \$20,000 or more. For this pair of questions (QI20N and QI22, corresponding to edited variables PINC1 and FINC1), the second question in the pair applied to the entire family. For answering these questions about total personal or family income, respondents were shown a list of other sources of income that they had not been previously asked about (e.g., child support payments). Respondents were instructed in answering these questions about total income to consider both these other sources of income and the sources of income they had been asked about previously.

If respondents had no other family members living in their respective dwelling units, FINC1 was assigned a legitimate skip code (Section 2.4.2). QI22 also was skipped and FINC1 was assigned a legitimate skip code if respondents reported in QI20N that they had personal incomes of \$20,000 or more. A third binary total family income variable, FAMINC1, was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household.

12.2.2 Finer Category Income Variables

Respondents also were asked to identify, both for themselves and for their families, finer categories of income within the two general categories of “\$20,000 or more” or “Less than \$20,000.” If respondents answered binary total income questions as “Less than \$20,000,” they were asked to report a finer category of income from incomes less than \$1,000 (including a net loss of income) up to \$20,000 in increments of \$1,000. Similarly, respondents who answered binary total income questions as “\$20,000 or more” were asked to report a finer category of income from \$20,000 up to \$50,000 in increments of \$5,000. If respondents’ or their families’ incomes were greater than \$50,000, they could select the following additional categories: “\$50,000 to \$74,999,” “\$75,000 to \$99,999,” “\$100,000 to \$149,999,” or “\$150,000 or more.”¹⁷⁵

Therefore, income variables for 2023 were created based on a maximum income category of \$150,000 or more. The variable PRINC3 for finer categories of personal income was created from responses to questions QI21A (for personal incomes of less than \$20,000) and QI21B (for personal incomes of \$20,000 or more). The variable FMINC3 for finer categories of family income was created from questions QI23A (for family incomes of less than \$20,000) and QI23B (for family incomes of \$20,000 or more). Thus, income categories in PRINC3 and FMINC3 ranged from 1 (less than \$1,000) to 30 (\$150,000 or more).

Questions QI23A or QI23B were skipped if the respondent had no other family members in the household. Therefore, when the imputed household roster variable IRFAMSKP indicated that no other family members were living in the household, FMINC3 was assigned a legitimate skip code of 99. Implementation of this edit included situations in which data existed in questions QI23A or QI23B but IRFAMSKP subsequently indicated that no other family members lived in the household.

A third total family income variable with finer categories of income, FMLYINC3, also was created. If other family members were living in the household, FMLYINC3 was set to be equal to the response from the mutually exclusive pair of total family income questions QI23A and QI23B. If no other family members lived in the household, FMLYINC3 was set to be equal to the responses from questions QI21A or QI21B. Thus, unlike FMINC3, FMLYINC3 was not assigned legitimate skip codes if no other family members lived in the household. Respondents who reported in question QI21B that their total personal income was the maximum income category (i.e., \$150,000 or more in 2023) were not asked question QI23B about their total family income. FMLYINC3 was assigned a code of 30 (\$150,000 or more) in this situation.

¹⁷⁵ The maximum income categories were “\$75,000 or more” prior to 2004 and “\$100,000 or more” in 2004-2014.

12.2.3 Additional Edits to Income Variables for Inconsistencies or Logical Inferences

Relatively little additional editing was done to the income variables, aside from assigning legitimate skip codes (Section 2.4.2). This section describes the additional edits that were implemented because of inconsistencies in the data or to make logical inferences.

As noted in Section 4.2.3, information was entered at the beginning of the interview about the state where the dwelling unit was located. This information was used in the income section to fill in the state-specific names for cash assistance from state or county welfare programs in question QI08N.¹⁷⁶ If the edited variable STATELOC for the state where the dwelling unit was located had been set to bad data because the state information entered by the interviewer or respondent was inconsistent with the sample data, any nonblank values in the edited variable corresponding to question QI08N (FAMPMT) were set to bad data. This edit was implemented because the incorrect state residence information could result in respondents not being asked about cash assistance from the program in the states where they actually lived. In turn, if FAMPMT had been set to bad data because STATELOC had been set to bad data, and FAMSVC did not indicate welfare assistance, then nonblank values in WELMOS also were set to bad data. In 2023, these edits were implemented for approximately 115 interviews, of which approximately 65 percent were web interviews.

Two types of edits were implemented for overall family income variables based on data about personal income:

- If the binary family income variable FINC1 indicated that the family income was less than \$20,000, and the corresponding personal income variable PINC1 had a missing value, then the respondent's personal income (PINC1) was logically inferred to be less than \$20,000. Similarly, if the binary personal income variable PINC1 was \$20,000 or more, and the family income variable FINC1 was skipped, then the total family income variable FAMINC1 indicated that the family income was \$20,000 or more.
- If the finer category for total personal income (PRINC3) was greater than the corresponding income category in FMINC3, then FMINC3 and FMLYINC3 were set to bad data. For example, if PRINC3 indicated that the respondent had an income of \$25,000 to \$29,999, and the respondent reported a total family income of only \$20,000 to \$24,999, then FMINC3 and FMLYINC3 were set to bad data. In 2023, these edits were implemented for approximately 360 interviews, of which approximately 70 percent were web interviews.

In addition, the recoded variables PINC2, FINC2, and FAMINC2 have been created since 2016 to facilitate comparison of the income data with data from 2004 to 2014, when the maximum income category was \$100,000 or more. These recodes were created by collapsing reported

¹⁷⁶ The names for state-specific welfare program names are updated annually as needed to ensure respondents provide accurate information on the receipt of cash assistance from these programs. Comparability of estimates across years could be affected if program names were *not* updated as needed.

incomes of \$150,000 or more with incomes of \$100,000 to \$149,999 to create a maximum category of 29 (\$100,000 or more).

12.3 Editing for Cell Phone Variables

Respondents were asked whether anyone in the household had a cell phone that was currently working (CELL2) and whether there was at least one phone in the household that was not a cell phone (CELL1). Minimal processing of data was done to these variables for cell phone coverage. Specifically, the unedited variables were assigned final variable names. Question CELL2 was named as CELLWRKNG and question CELL1 was named as CELLNOTCL for the edited variables.

12.4 Imputation for the Income Variables

The imputation of income was separated into two imputation sets. The first set involved the imputation of all the binary income variables (i.e., “yes/no” questions about the following sources of income: social security, supplemental security income, welfare cash assistance, welfare noncash assistance, wages, and SNAP benefits), the number of months on welfare (the only variable that was not binary in this imputation set), and a yes/no question regarding whether the respondent’s income or the respondent’s household family income was \$20,000 or more (including income from the sources referenced in the previous questions). This first set was processed using the single response propensity (RP)/multiple prediction (PRD) PMN type described in Section 3.4.3. The second imputation set for finer categories of income consisted of imputing more specific income categories for the respondent and the respondent’s family in the household. This set was processed using the single RP/single PRD PMN type as described in Section 3.4.1.

The income variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation.

Imputations for all income variables were conducted separately within four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. The segregation into age groups was done to exploit the correlation between the income variables and age and to allow parallel processing of the data (thus reducing the time it takes to implement the procedures). Imputation procedures were implemented for the combined data from all four quarters, and the quarter in which the interview took place (i.e., Quarters 1, 2, 3, or 4) was included as a covariate in the RP and PRD models for PMN. See Chapter 3 for more information.

All income RP and PRD models used a state-rank variable as a covariate. States were ranked by the weighted proportion of respondents whose personal incomes during the prior calendar year were greater than or equal to \$20,000. See Section 3.8.2 for a general discussion of state-rank variables.

12.4.1 Imputation for Binary Income Variables (Imputation Set 1)

The PMN imputation type used for the binary income variables was single RP/multiple PRD. The RP model is described in Section 12.4.1.1. The PRD model for the first binary income variable imputed each year, family social security (FAMSOC), is described in Section 12.4.1.2, and the provisional hot-deck step for FAMSOC is described in Section 12.4.1.3. The remaining PRD and provisional hot-deck steps described in Section 12.4.1.4 list deviations from the analogous steps for family social security. The final hot-deck step applied to all binary income variables is described in Section 12.4.1.5. Finally, a recode for the GOVTPROG variable, made from four imputation-revised binary income variables, is described in Section 12.4.1.6.

For the binary income models that predict whether a respondent had a given source of income, other sources of income were useful covariates. Therefore, provisionally imputed values were used as covariates in subsequent models within the set. The order in which missing values for the binary income variables were provisionally imputed is listed in [Table 12.2](#).

Table 12.2 Imputation Order for Binary Income Variables (Imputation Set 1)

Income Type	Edited Variable Names	Imputation-Revised Variable Names
Family Social Security	FAMSOC	IRFAMSOC
Family Supplemental Security Income	FAMSSI	IRFAMSSI
Family Welfare Payments	FAMPMT	IRFAMPMT
Family Other Welfare Services	FAMSVC	IRFAMSVC
Family SNAP Benefits	FSTAMP	IRFSTAMP
Family Months on Welfare	WELMOS	IRWELMOS
Total Personal/Family Income	PINC1, FINC1, FAMINC1	IRPINC1, IRFINC1, IRFAMIN1

12.4.1.1 Response Propensity Step

The RP models for the binary income variables utilized the preliminary analysis weight. All respondents were in the domain for the binary income variables (i.e., eligible to have a valid value for these variables). For the single RP model, a domain member was considered an item respondent if and only if all of the variables listed in [Table 12.2](#) were nonmissing. See Appendix D for details on the covariates used in the RP models for these variables.

12.4.1.2 First Prediction Step (FAMSOC)

FAMSOC is a binary variable; after imputation, its only values are “yes” and “no” since all other missing data codes are replaced with imputed values. Using the adjusted weights that are outputs of the RP step, FAMSOC was modeled using logistic regression and the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean used in later hot-deck steps was the predicted probability that the respondent’s family-in-household received income from social security in the preceding calendar year.

12.4.1.3 First Provisional Hot-Deck Step (FAMSOC)

The provisional hot-deck step applied to the FAMSOC variable is a simplified version of the final hot-deck step for the binary income variables. The only purpose is to fill in missing values so that FAMSOC can be used as a covariate in the PRD models for binary income variables imputed later in the sequence. Section 3.4.2 describes the concept of a provisional hot-deck step in more detail. The final hot-deck step is described in Section 12.4.1.5 and in tabular form in Appendix E. This provisional hot-deck step included the following:

- The predictive mean vector included only one element: the predicted mean from the preceding step.
- No logical constraints were used.
- The following likeness constraints were used:
 - IRFAMSKP of donor = IRFAMSKP of recipient. This is a likeness constraint in this hot-deck step, not a logical constraint. The creation of IRFAMSKP is described in Section 11.3.2.
 - Donor's predicted mean must be within 5 percent of recipient's predicted mean (delta constraint).
 - Donor must match recipient with respect to whether there are adults aged 65 or older in the household. The variable used in this constraint was IRHH65, whose creation is described in Section 11.3.2.
- In the first attempt to find a donor, all three likeness constraints were applied. In the second attempt, the delta constraint was removed.

The provisionally imputed version of FAMSOC was called INTFAMSOC. INTFAMSOC was used as a covariate in the rest of the binary income PRD models.

12.4.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Binary Income Variables (FAMSSI, FAMPMT, FAMSVC, FSTAMP, WELMOS, and FAMINC1)

PRD models were fit for FAMSSI, FAMPMT, FAMSVC, FSTAMP, and FAMINC1¹⁷⁷ in the same manner as for FAMSOC, as described previously. Only the PRD model for WELMOS was different. The domain of the WELMOS model included only respondents who reported that their family-in-household received welfare payments and/or other welfare services during the preceding calendar year, as defined by FAMPMT and FAMSVC. Least squares regression (not logistic regression) was used, where the dependent variable was a standard logit,¹⁷⁸ such that $Y = \text{logit}(p)$ and p = number of months on welfare divided by 12. The REGRESS procedure in SAS-callable SUDAAN was used to fit the model. The predicted mean from the WELMOS model was the predicted probability of receiving welfare in a given month in the previous calendar

¹⁷⁷ The response variable for the model for total income was total family income (not total personal income), and the provisionally imputed version of WELMOS was not included as a covariate in the model.

¹⁷⁸ The Cox empirical logit was used when an individual was on welfare for all 12 months (Cox & Snell, 1989).

year, given that the respondent received welfare payments and/or welfare services in the previous calendar year.

The provisional imputation steps for the other variables were implemented in the same manner as for FAMSOC, as described previously, except that the likeness constraints sometimes differed slightly. The following deviations are noted:

- Only the provisional hot-deck step for FAMSOC included the constraint involving the IRHH65 variable.
- For FAMPMT and FAMSVC, the following likeness constraint was used: donor must match recipient with respect to whether there are children younger than 18 in the household. The variable used in this constraint was IRKID17, whose creation is described in Section 11.3.2. This likeness constraint was used in both the first and second attempts to find a donor.
- FAMINC1 is imputed in the final hot-deck step because it is the last variable in the set. See Section 12.4.1.5 for details on the final hot-deck step.

12.4.1.5 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the binary income variables' final hot-deck step are available in Appendix E. This section explains the more general ideas behind the hot-deck step applied to these variables.

Because nine imputation-revised variables are created in this step, and almost any subset of them can be missing, there are 319 missingness patterns for the binary income variables. For the first 127 patterns, WELMOS is nonmissing; these are summarized as pattern 5 in Appendix E. For the remaining 192 patterns, WELMOS is missing; these are summarized as patterns 1 through 4 in Appendix E.

- When WELMOS is nonmissing, any subset of FAMSOC, FAMSSI, FAMPMT, FAMSVC, FSTAMP, PINC1, and FINC1 may be missing. However, if none of them are missing, then nothing is missing and no imputation is required. This leads to $2^7 - 1 = 127$ missingness patterns.¹⁷⁹
- When WELMOS is missing, the elements of the predictive mean vector related to FAMPMT, FAMSVC, and WELMOS depend on the values of FAMPMT and FAMSVC. There are six relevant combinations of FAMPMT and FAMSVC (Table 12.3). For each of the six, any subset of FAMSOC, FAMSSI, FSTAMP, PINC1, and FINC1 may be missing. This leads to $6 \times 2^5 = 192$ missingness patterns.
 - For the combinations labeled 1 in Table 12.3, the recipient is known to have received welfare payments and/or welfare services and neither is missing. Here, no predictive

¹⁷⁹ Technically, if FAMPMT and FAMSVC were both missing, then WELMOS was necessarily missing. Therefore, $2^6 = 64$ out of these 127 missingness patterns are logically impossible and need not be set up.

mean vector elements are required for FAMPMT and FAMSVC,¹⁸⁰ and there is no need to manipulate the element for WELMOS, which is already conditional on receipt of welfare as defined by FAMPMT and FAMSVC.

- For combinations 2 and 3, the recipient is also known to have received welfare payments or welfare services, but one is missing. Here, a predictive mean vector element (i.e., a predicted mean) is required for the missing one of the two, but again there is no need to manipulate the corresponding element for WELMOS.
- For combination 4, the recipient is known not to have received welfare payments, but welfare services is missing. Here, a predictive mean vector element is required for FAMSVC. The WELMOS predicted mean must be made conditional on the receipt of welfare services in the previous calendar year. Combination 5 is similar, but FAMPMT is the missing one.
- For combination 6, both FAMPMT and FAMSVC are missing. Here, predictive mean vector elements are required for both. The WELMOS predicted mean must be made conditional on the receipt of welfare payments and/or welfare services. This probability is unknown but can be crudely approximated by assuming the two are independent. Under that assumption, the probability of receiving either or both is one minus the probability of receiving neither, as expressed by $(1 - \text{PMT}) \times (1 - \text{SVC})$.¹⁸¹
- If FAMPMT = “No” and FAMSVC = “No,” then WELMOS is assigned a skip code and is therefore nonmissing. This combination is irrelevant for the purposes of this discussion, which only covers situations where WELMOS is missing.
- Combinations 1 through 3 in [Table 12.3](#) correspond to pattern 1 in Appendix E, combination 4 corresponds to pattern 2, combination 5 corresponds to pattern 3, and combination 6 corresponds to pattern 4.

Table 12.3 Cross-Classification of FAMPMT and FAMSVC; Relevant Combinations for Binary Income Predictive Mean Vector when WELMOS Is Missing

FAMPMT	FAMSVC		
	Yes	No	Missing
Yes	1	1	2
No	1	N/A	4
Missing	3	5	6

N/A = not applicable.

¹⁸⁰ In general, in either the single RP/multiple PRD or multiple RP/multiple PRD types of PMN, only the elements of the predictive mean vector corresponding to missing responses were used (see Chapter 3).

¹⁸¹ A reasonable alternative method that requires no assumption of independence would be to model FAMPMT and FAMSVC together as a categorical variable with four levels: both, only FAMPMT, only FAMSVC, and neither. Then, the probability of receiving welfare payments and/or other welfare services is simply the sum of the first three predicted means. The assumption of independence is certainly questionable.

Both the likeness and logical constraints are written in such a way that they can be summarized using the five broad missingness patterns listed in Appendix E. For example, one likeness constraint states, “Donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing).” This constraint does not apply to the first 127 missingness patterns, nor does it apply to the 64 missingness patterns in combination 6 where both FAMPMT and FAMSVF are missing. However, instead of carefully listing the exact missingness patterns to which each constraint applied, the constraints were written using conditions that allowed them to be applied only where necessary.

The few logical constraints are due to relationships between (1) FAMPMT, FAMSVF, and WELMOS and (2) PINC1, FINC1, and FAMINC1. A key likeness constraint exploited the high degree of association among welfare payments, welfare services, SNAP benefits, binary total income (at the personal and family level), and months on welfare. If a recipient required imputation for one or more of these six variables, but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables.

As stated in the introduction to Section 12.4, the variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation. The constraints are fairly complex and extensive, especially when several of the other dichotomous income variables are missing. Also, when several variables are missing, it is difficult to find a donor on the first try because the donor must be close to the recipient for all predicted means corresponding to the missing variables.

12.4.1.6 Recodes for Additional Analyses

A dichotomous recoded income variable GOVTPROG indicated whether the respondent participated in any government assistance programs. It was created from four imputation-revised variables: family supplemental security income (IRFAMSSI), family SNAP benefits (IRFSTAMP), family welfare payments (IRFAMPMT), and family welfare services (IRFAMSVF). Although a variety of recoded variables were created, only GOVTPROG is described here because it was used as a covariate in subsequent health insurance imputation models and mental health models. See Section 13.3 for details on the imputation of missing values in the health insurance variables.

12.4.2 Imputation for Finer Category Income Variables (Imputation Set 2)

Three income variables resulted from editing the questions in the finer category income phase: personal total income (PRINC3), total family income if there are other family members in the household (FMINC3), and total family income (FMLYINC3). These edited variables are described in Section 12.2.2. All three imputation-revised variables derived from their edited counterparts were created using the single RP/single PRD PMN type described in Section 3.4.1. The single PRD model was fit for total family income (FMLYINC3), but the item nonrespondent also received values from the donor for PRINC3 and FMINC3 if those were missing. There were no deviations from this general approach.

12.4.2.1 Response Propensity Step

The RP models for the finer category income variables utilized the preliminary analysis weight. All respondents were in the domain for this imputation set. Item respondents were those with nonmissing values for PINC2 and FINC2.¹⁸² See Appendix D for details on the covariates used in the RP models for these variables.

12.4.2.2 Prediction Step

Each of the three finer category income variables was ordinal with 30 levels. Each response category covers an interval of income, with levels as follows:

- 1 LESS THAN \$1,000
- 2 \$1,000 - \$1,999
- 3 \$2,000 - \$2,999
- 4 \$3,000 - \$3,999
- 5 \$4,000 - \$4,999
- 6 \$5,000 - \$5,999
- 7 \$6,000 - \$6,999
- 8 \$7,000 - \$7,999
- 9 \$8,000 - \$8,999
- 10 \$9,000 - \$9,999
- 11 \$10,000 - \$10,999
- 12 \$11,000 - \$11,999
- 13 \$12,000 - \$12,999
- 14 \$13,000 - \$13,999
- 15 \$14,000 - \$14,999
- 16 \$15,000 - \$15,999
- 17 \$16,000 - \$16,999
- 18 \$17,000 - \$17,999
- 19 \$18,000 - \$18,999
- 20 \$19,000 - \$19,999
- 21 \$20,000 - \$24,999
- 22 \$25,000 - \$29,999
- 23 \$30,000 - \$34,999
- 24 \$35,000 - \$39,999
- 25 \$40,000 - \$44,999
- 26 \$45,000 - \$49,999
- 27 \$50,000 - \$74,999
- 28 \$75,000 - \$99,999
- 29 \$100,000 - \$149,999
- 30 \$150,000 OR MORE

¹⁸² See Section 12.2.3 for a description of PINC2 and FINC2.

The FMLYINC3 variable was modeled using the LIFEREG procedure in SAS/STAT® software.¹⁸³ This procedure was used for regression modeling of continuous nonnegative random variables, such as survival times and income, by fitting models that are sometimes referred to as “failure time models.” This particular type of model, which was assumed for the response variable representing income, can be written as

$$y = X\beta + \varepsilon,$$

where y is a vector of observed responses, X is the matrix of covariates, β is the parameter vector, and ε is a vector of error terms. Although the underlying random variable y is assumed to be continuous, the LIFEREG procedure allows the variable to be reported in interval categories, consistent with the 30 NSDUH income intervals defined previously for these finer category income variables. The contribution of an individual with covariates in the matrix X to the overall likelihood is simply the probability mass assigned by the model to the interval (l, u) containing the actual continuous income for that individual. For this interval, l represents the lower bound and u represents the upper bound. This contribution has the form $F(u|X, \beta, \sigma^2) - F(l|X, \beta, \sigma^2)$, where F is a cumulative distribution function and σ^2 represents the variance of the individual responses. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.

LIFEREG allows several choices for the functional form of the parametric model that correspond to the error distribution, including the two-parameter log-logistic, lognormal, and Weibull and the three-parameter generalized gamma. Compared with the other models, the gamma distribution (DISTRIBUTION=GAMMA) provided a better overall fit, as indicated by likelihood techniques, and is used for modeling finer categories of income.

The predicted mean used in the subsequent hot-deck step was the term $X\beta$, which was the predictive mean value. This value was a monotonic function of the conditional mean of the modeled income distribution at a given individual set of values of the regression covariates. Specifically, $X\beta$ was a translation of the estimated mean of log income.

12.4.2.3 Hot-Deck Step

The hot-deck step for the finer category income variables is an example of univariate matching, multivariate assignment (see [Table 3.3](#)). The only predicted mean used in the hot-deck step was related to FMLYINC3, but the recipient received values for PRINC3 and FMINC3 if either or both were missing. The imputation-revised version of FMLYINC3 (IRFMLYIN3) was created from the imputation-revised versions of PRINC3 and FMINC3, called IRPRINC3 and IRFMINC3, respectively. The constraints ensure consistency with existing information such as IRPINC1 and IRFINC1 from the preceding imputation set, and PRINC3 and FMINC3 if nonmissing.

The finer category income variables generally have among the lowest response rates of any NSDUH variables that undergo imputation. As a result of the absence of many constraints, the

¹⁸³ Details about the LIFEREG procedure are discussed in the *SAS/STAT® 13.2 User's Guide: High-Performance Procedures* (SAS Institute Inc., 2014).

single predicted mean, and the large domain, most respondents are handled in the first attempt to find a donor.

12.4.2.4 Recodes for Additional Analyses

The recoded variable INCOME classified the families of respondents into four income levels: less than \$20,000; \$20,000 to \$49,999; \$50,000 to \$74,999; and \$75,000 or more. Another recoded variable (INCOME5) was created to take advantage of an extra level of income. This variable had five levels: the first three levels were equivalent to INCOME, but the last level of INCOME was separated into two levels: \$75,000 to \$99,999 and \$100,000 or more. Both INCOME and INCOME5 were recodes of the variable IRFMLYIN3. A variety of recoded variables were created but are not discussed in this report. However, as with GOVTPROG, the variable INCOME is discussed here because it was used as a covariate in subsequent health insurance models and mental health models. See Section 10.8 for details on the imputation of missing values in the mental health variables, and see Section 13.3 for details on the imputation of missing values in the health insurance variables. INCOME5 also is discussed because it is similar to the INCOME variable.

The recoded variables IRPINC2, IRFINC2, and IRFAMIN2 were created to be comparable with the variables of the same names created in 2004-2014, in which the finer categories of income had 29 levels with the top level being "\$100,000 OR MORE." These recodes were made by collapsing the top two levels of variables IRPRINC3, IRFMINC3, and IRFMLYIN3.

13. Editing and Imputation for the NSDUH Health Insurance Variables

13.1 Introduction

This chapter discusses procedures for logically editing and imputing data from the section pertaining to health insurance coverage. The health insurance section is interviewer administered for in-person respondents and self-administered for web respondents. Respondents (or other household members serving as proxies) were asked whether they were currently covered by different types of health insurance. If private health insurance coverage was reported, respondents were asked whether that included coverage for substance abuse treatment or mental health services. Data also were collected on periods when respondents never had health insurance coverage, most recent coverage (if currently without health insurance), and reasons for losing health insurance coverage or for never having had coverage (if applicable).

Consistent with most of the imputation-revised variables that were discussed in the previous chapters of this report, imputations for the health insurance variables in the 2023 National Survey on Drug Use and Health (NSDUH) were performed using the predictive mean neighborhood (PMN) methodology detailed in Chapter 3. This method has been used since the 2002 NSDUH to impute for specific health insurance variables (e.g., Medicare coverage) using all relevant questions in the questionnaire.

Imputation-revised variables are created for the following five sources of health insurance coverage: (1) Medicare; (2) Medicaid or the Children's Health Insurance Program (CHIP); (3) coverage through the United States military,¹⁸⁴ such as the Civilian Health and Medical Program of the Uniformed Services (CHAMPUS); (4) private health insurance, such as insurance obtained through work; and (5) any other kind of health insurance coverage. In addition to these five imputation-revised variables, the health insurance imputation process creates the overall health insurance variable IRINSUR4. This variable combines information from the detailed imputation-revised variables into one summary measure of coverage.

13.2 Editing the Health Insurance Variables

An important aspect of editing the health insurance variables consisted of assigning legitimate skip codes based on the skip logic in the health insurance section (Section 2.4.2). [Figure 13.1](#) shows the logic for asking questions in the health insurance section. For example, if respondents answered "no" (where applicable) to questions QHI01 through QHI06 and then reported in QHI11 that they were not currently covered by any kind of health insurance (QHI11 = 2), legitimate skip codes were assigned to HLCNOTYR (anytime in the past 12 months that respondents were without health insurance, corresponding to question QHI13) and HLCNOTMO

¹⁸⁴ Because people who are serving on active duty in the United States military are ineligible for NSDUH, coverage through the United States military applies to respondents who are not on active duty, such as dependents of military personnel or military retirees.

(number of months that respondents were without health insurance in the past 12 months, corresponding to question QHI14). Similarly, if respondents reported some type of current health insurance coverage in QHI01 through QHI06 or in QHI11, then edited variables corresponding to questions QHI15 through QHI18 were assigned legitimate skip codes (i.e., HLCLAST through HLNVSOR).

Question QHI18 (reasons that the respondent never had health insurance) was an “enter all that apply” question. Therefore, the edited variables corresponding to question QHI18 (HLNVCOST through HLNVSOR) were assigned a code of 1 (Response entered) if the corresponding response category was chosen from QHI18 and a code of 6 (Response not entered) if the corresponding response category was not chosen but at least one response had been entered in QHI18 (Section 2.4.4).

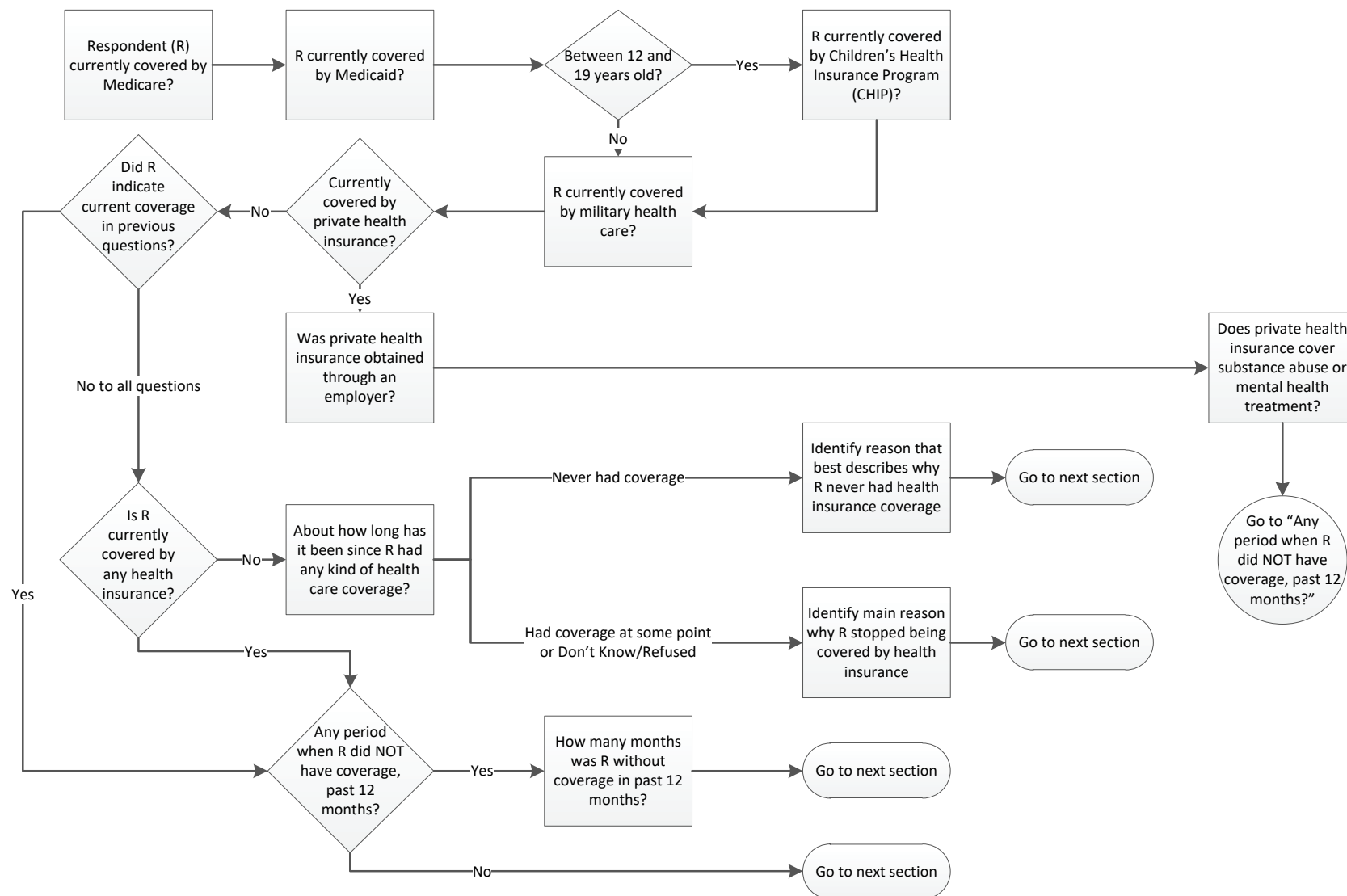
13.2.1 Editing Because of Incorrect State Information

Question QHI02 asked whether respondents were currently covered by Medicaid, and question QHI02A asked whether respondents aged 12 to 19 were covered by CHIP. The computer-assisted interviewing (CAI) program filled in names for state-specific Medicaid program or CHIP names to aid respondent identification.

If the information entered at the beginning of the interview by the interviewer or respondent about the state where the dwelling unit was located (edited variable STATELOC) had been set to bad data because of inconsistencies with sample data (Section 4.2.3), then the edited variable CAIDCHIP for coverage by Medicaid or CHIP (see Section 13.2.2) usually was assigned a bad data code as well. The rationale for this edit was that the name for the state’s Medicaid program in question QHI02 or CHIP in question QHI02A was governed by the state that was reported by the interviewer or respondent. If the state information that the interviewer or respondent reported was inconsistent with sample data, then the answers to QHI02 or QHI02A were deemed questionable.

An exception to this assignment of bad data codes concerned the special situation in which respondents were routed to questions QHI15 (time since the respondent last had health insurance) and QHI17 (main reason for losing health insurance coverage), because being routed to QHI15 and QHI17 indicated that these respondents did not currently have health insurance coverage. In this situation, CAIDCHIP retained a code of 2 (i.e., “no”), even if STATELOC had been set to bad data, for consistency with information from QHI15 or QHI17 that these respondents were not currently covered by any type of health insurance.

Figure 13.1 Routing Logic in the Health Insurance Section of the Questionnaire



NOTE: Respondents immediately go to the next section (i.e., without further health insurance questions being asked) if current health insurance coverage status is unknown.

13.2.2 Determination of Any Health Insurance Coverage

Determining whether respondents were currently covered by any type of health insurance was critical for determining which subsequent questions applied to individual respondents. Subsequent questions asked about periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, or reasons for never having health insurance. The following content issues are relevant for determining whether respondents had health insurance coverage:

- As noted previously, respondents who were aged 12 to 19 are asked in question QHI02A whether they were covered by CHIP. Government experts in the health insurance field advised the Substance Abuse and Mental Health Services Administration that it would be virtually impossible to produce separate estimates of coverage by the Medicaid program (question QHI02) and coverage by CHIP. For this reason, the variable CAIDCHIP was created from responses to questions QHI02 and QHI02A.
- The state-specific Medicaid or CHIP names that are used to customize questions QHI02 and QHI02A may be updated for the CAI program across years to reflect any changes to these names at the state level. This helps ensure that respondents provide accurate information on coverage by these programs.
- If STATELOC had a valid value (Sections 4.2.3 and 13.2.1), CAIDCHIP was assigned a code of 1 (yes) if an affirmative response occurred in either QHI02 or QHI02A (if applicable). CAIDCHIP was coded as 2 (no) if QHI02 was answered as “no” and (a) QHI02A also was answered as “no” (for respondents who were aged 12 to 19) or (b) QHI02A had been legitimately skipped (for respondents aged 20 or older). Otherwise, CAIDCHIP was coded as 97 (refused) if a code of 97 occurred in either QHI02 or QHI02A, or 94 (don’t know) if a code of 94 (and no code of 97) occurred in these items. Remaining situations that did not meet any of these criteria were coded as 98 (blank).
- Respondents who answered “no” to all questions about Medicare, Medicaid, CHIP (if applicable), military health coverage, and private health insurance were asked a follow-up question (QHI11) to determine if they were covered by any type of health insurance. Responses to this question were used to determine subsequent routing in the health insurance section depending on whether respondents currently had or did not have health insurance. The variable HLTINNOS was created from QHI11.
- A recoded “any health insurance” variable, ANYHLTIN, was created from responses to MEDICARE (from QHI01), CAIDCHIP (from QHI02 and QHI02A), CHAMPUS (from QHI03), PRVHLTIN (from QHI06), and HLTINNOS (from QHI11).
 - If any affirmative response was reported in any of the above variables, ANYHLTIN was coded as 1 (yes).
 - Otherwise, if HLTINNOS (from QHI11) had been answered as “no” (and by definition, preceding questions had been answered as “no”), ANYHLTIN was coded as 2 (no).
 - If ANYHLTIN was not already coded as 1 or 2, it was coded as 97 (refused) or 94 (don’t know), as follows: (a) if a code of 97 (refused) occurred in any of the above

- health insurance items, ANYHLTIN was coded as 97; or (b) ANYHLTIN was coded as 94 (don't know) if a code of 94 (but no code of 97) occurred in the above items.
- For remaining situations (e.g., if variables had been set to bad data, or a break-off had occurred), ANYHLTIN retained a code of 98 (OTHER MISSING).

In addition, because the names for Medicare and Medicaid sound similar, follow-up questions were included in the health insurance section of the interview for respondents who were (1) younger than 65 and reported being covered by Medicare (which is for people aged 65 or older or certain disabled individuals) or (2) aged 65 or older and reported being covered by Medicaid (which is for low-income individuals or disabled individuals). In these situations, respondents were asked to confirm their previous answer. If respondents indicated on follow-up that their previous answer to Medicare or Medicaid coverage was incorrect, it was determined that they were not covered by that particular type of health insurance. However, if respondents did not know or refused to confirm whether their previous answer to the Medicare or Medicaid coverage question was correct, the editing procedures continued to classify them as having that type of coverage. Stated another way, the only response on follow-up that overruled a previous report of coverage by Medicare or Medicaid was for respondents to indicate that they did not have that type of coverage.

13.2.3 Editing of Health Insurance Follow-Up Variables Based on Current Coverage

Information from ANYHLTIN and a preliminary edited variable SKHLCCOV¹⁸⁵ were used to determine whether variables pertaining to periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, and reasons for never having health insurance should be assigned legitimate skip codes (Section 2.4.2).

- If respondents reported some health insurance coverage in QHI01 to QHI11, legitimate skip codes were assigned to the edited variables corresponding to questions QHI15 to QHI18. Questions QHI15 to QHI18 applied to respondents who currently were without coverage and pertained to situations in which respondents formerly or never had coverage.

¹⁸⁵ SKHLCCOV was a precursor to the final variable ANYHLTIN but was used in some places in the editing program to determine whether respondents were legitimately skipped out of certain follow-up questions about health insurance coverage. Both variables indicate whether respondents currently had health insurance coverage. SKHLCCOV was coded as 3 if respondents reported in question QHI02A that they were covered by CHIP but they did not report coverage by Medicaid, Medicare, military health insurance (e.g., CHAMPUS), or private health insurance. These values of 3 in SKHLCCOV were recoded to 1 in ANYHLTIN.

- If respondents currently were without any health insurance coverage (i.e., applicable questions in QHI01 to QHI11 all were answered as “no”), legitimate skip codes were assigned to the edited variables corresponding to questions QHI13 and QHI14. Questions QHI13 and QHI14 applied to respondents who currently had coverage and pertained to periods when these respondents were without coverage in the past 12 months.
- If it was unknown from questions QHI01 to QHI11 whether respondents currently had health insurance coverage, the CAI program skipped respondents out of all remaining health insurance questions. The edited variables corresponding to questions QHI13 to QHI18 had missing values.

13.2.4 Miscellaneous Issues for Editing the Health Insurance Variables

[Table 13.1](#) discusses edit specifications for the health insurance variables. For example, the data could indicate that respondents were covered currently by Medicare, Medicaid, CHIP (for respondents who were aged 12 to 19), some type of military health coverage (e.g., CHAMPUS or the VA), or private health insurance. If respondents were reported to have been currently covered by all of the types of insurance they were asked about, a flag was set and included on the data file. The original data were retained, but this flag was designed to alert analysts to the presence of this unlikely data pattern.^{[186](#)}

Table 13.1 Edits Pertaining to the Health Insurance Section

Issue	Edits Implemented
The respondent (R) reported being currently covered by Medicare, Medicaid, Children’s Health Insurance Program (CHIP) (if aged 12 to 19), military coverage, and private health insurance.	A flag (HLCALLFG) was provided to indicate that this pattern occurred, but no further editing was done to the data.
The R reported being currently covered by Medicare, Medicaid, military coverage, and private health insurance, the only types of current coverage that were asked about in 1999.	A flag (HLCALL99) was provided to indicate that this pattern occurred, but no further editing was done to the responses. This HLCALL99 variable was created in the same manner as the HLCALLFG variable in 1999.
The R’s only indication of current health insurance coverage came from reports of coverage by Medicaid or CHIP, but the state location variable STATELOC had been set to bad data.	Nonblank values in the variables pertaining to any period in the past 12 months when the R was without health insurance (HLCNOTYR, corresponding to question QHI13) and the number of months that the R was without health insurance in the past 12 months (HLCNOTMO, corresponding to question QHI14) were replaced with bad data codes. This edit was not done if the R indicated current coverage by Medicare, the military, or private health insurance.

¹⁸⁶ A similar flag was set if respondents reported that they were covered currently by all four types of health insurance that were asked about in 1999 (Medicare, Medicaid, military health coverage, private health insurance), even if they did not report being covered by CHIP (if aged 12 to 19).

Table 13.1 Edits Pertaining to the Health Insurance Section (continued)

Issue	Edits Implemented
The R had some indication of current coverage from at least one of the five sources of insurance listed above. However, the R also was reported to have had a period in the past 12 months when the R was without health insurance. Further, the R was reported to have been without health insurance for 12 of those months.	No editing was done when this pattern occurred. The rationale for not doing any editing was that the R may just recently have gotten insurance or have become qualified for insurance.
The R had no indication of current coverage from any of the five sources of insurance listed above. If the R (or proxy) answered “don’t know” or “refused” when asked when the R last had coverage, the R was routed to questions about what coverage the R last had, and why the R lost health insurance coverage. That is, the skip logic assumed that the R had some prior history of coverage, but that may not necessarily have been the case.	If the R was reported to have had previously some form of health insurance or medical coverage, or if some reason was given why the R lost insurance coverage, legitimate skip codes were assigned to the variables pertaining to reasons that the R never had coverage. That is, the implicit assumption made in the computer-assisted interviewing (CAI) skip logic was verified by an answer indicating some prior history of health insurance coverage. However, if nothing was reported to indicate that the R previously had health insurance, the skipped variables pertaining to reasons for never having had insurance retained codes of blank.
The R was male but reported in QHI17 that the reason for losing health insurance coverage was because of eligibility for Medicaid or medical insurance only while pregnant.	The edited variable HLLOSRSN (corresponding to QHI17) was set to bad data.

13.2.5 Editing Types of Health Insurance and Any Other Health Insurance Variables, Imputation Sets 1 and 2

As noted previously, the editing process for health insurance combined data from questions on whether respondents were covered by Medicaid or CHIP to create the variable CAIDCHIP (whether someone was covered by Medicaid or one of the state children’s health insurance programs). This CAIDCHIP variable and all other edited variables in [Table 13.2](#), except HLTINNOS, were used directly as base variables for imputation.

A respondent was routed to QHI11 (whether the respondent was covered by any kind of health insurance at the time of the survey) if they answered “no” to all the other health insurance questions. All other respondents were given a legitimate skip value to the variable HLTINNOS, as shown in [Table 13.2](#). Therefore, it was possible that the imputation-revised versions of the four specific health insurance variables would all have had a value of “no,” and the value of HLTINNOS would have been a legitimate skip, if one or more of the “no” values was imputed. In this instance, another variable was needed to reflect the fact that a respondent could have had a valid “yes” or “no” imputed value for any other health insurance, even though the respondent was never asked QHI11 and was assigned a legitimate skip code. The ANYOTHER variable was created using the preliminary edited variable, SKHLCCOV (Section 13.2.3), which indicated whether a respondent was covered by any health insurance. SKHLCCOV and ANYOTHER were defined as follows:

SKHLCCOV =

- 1 (or 3¹⁸⁷) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, PRVHLTIN = 1, or HLTINNOS = 1; else
- 2 if CAIDCHIP = 2, MEDICARE = 2, CHAMPUS = 2, PRVHLTIN = 2, and HLTINNOS = 2; else
- missing value code if the nonmissing values of CAIDCHIP, MEDICARE, CHAMPUS, PRVHLTIN, and HLTINNOS are all "2," and at least one of these variables had a missing response.

ANYOTHER =

- legitimate skip code (99) if any CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, or PRVHLTIN = 1; else
- SKHLCCOV.

Table 13.2 Mapping of Raw Health Insurance Variables to Edited Variables and Base Variables for Imputation

Question Number ¹	Question Text ²	Edited Variable ³	Used to Create CAIDCHIP	Used to Create MEDICARE	Used to Create CHAMPUS	Used to Create PRIVATE	Used to Create ANYOTHER
QHI01, QHI01v	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)	No	Yes	No	No	Yes
QHI02, QHI02v	Is the respondent covered by Medicaid or Medical Assistance?	N/A ⁴	Yes	No	No	No	Yes
QHI02A	Is the respondent currently covered by a Children's Health Insurance Program operated by your state of residence? ⁵	N/A ⁴	Yes	No	No	No	Yes

¹⁸⁷ See Section 13.2.3 for discussion of the assignment of a code of 3 to SKHLCCOV. Respondents with SKHLCCOV = 3 were treated in the same manner as those with SKHLCCOV = 1.

Table 13.2 Mapping of Raw Health Insurance Variables to Edited Variables and Base Variables for Imputation (continued)

Question Number ¹	Question Text ²	Edited Variable ³	Used to Create CAIDCHIP	Used to Create MEDICARE	Used to Create CHAMPUS	Used to Create PRIVATE	Used to Create ANYOTHER
QHI03	Is the respondent currently covered by TRICARE, CHAMPUS, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)	No	No	Yes	No	Yes
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)	No	No	No	Yes	Yes
QHI11	Is the respondent currently covered by any kind of health insurance, that is, any policy or program that provides or pays for medical care?	HLTINNOS (1 = yes, 2 = no, 99 = legitimate skip ⁶)	No	No	No	No	Yes

¹ The “v” questions were asked to verify the answer given in the previous question for respondents who were younger than 65 and a Medicare recipient or older than 65 and a Medicaid recipient.

² The questions provided in this table are abbreviated versions of those given in the questionnaire.

³ Missing values in these edited values were represented by standard missing value codes.

⁴ N/A = not applicable. Individual final edited variables were not created for Medicaid coverage or Children’s Health Insurance Program coverage for inclusion on NSDUH data files.

⁵ The questionnaire did not ask the question exactly in this way. It identified the specific program, depending upon the state of residence entered by the respondent and was asked only of respondents aged 12 to 19.

⁶ Respondents were assigned a legitimate skip for HLTINNOS if they answered “yes” or gave no answer to at least one of the other health insurance questions.

13.3 Imputation for the Health Insurance Variables

Missing values for the health insurance variables were replaced with valid values using the health insurance method.¹⁸⁸ Indicator variables for specific types of health insurance (e.g., coverage by Medicare) were divided into two imputation sets and were imputed using the PMN methodology (see Chapter 3):

- Imputation Set 1: IRMCDCHIP, IRMEDICR, IRCHMPUS, and IRPRVHLT (types of health insurance); and
- Imputation Set 2: IROTHHLT (any other health insurance).

The variables in imputation set 1 are indicators of coverage for Medicaid/CHIP, Medicare, CHAMPUS or similar coverage through the United States military, and private health insurance, respectively. These variables have item response rates around 97 percent. The variable IROTHHLT in imputation set 2 is an indicator of “any other health insurance.” This variable has a response rate of about 95 percent. Together, these five imputation-revised health insurance variables were used to create an indicator of “any health insurance” coverage, IRINSUR4, as discussed in Section 13.3.3.

13.3.1 Types of Health Insurance Variables, Imputation Set 1

The first four imputation-revised variables were created using the single response propensity (RP)/multiple prediction (PRD) PMN type. This approach was appropriate due to the high response rates for the variables being imputed in this set, and there were no noteworthy deviations from this general approach. The four imputation-revised variables created were yes/no indicators of whether the respondent had health insurance from Medicaid or a state children’s health insurance program (IRMCDCHIP), Medicare (IRMEDICR), CHAMPUS or other similar coverage for military personnel (IRCHMPUS), or private health insurance (IRPRVHLT). The health insurance indicators were imputed from base variables in the following order: CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN.

13.3.1.1 Response Propensity Step

The RP models for imputation set 1 utilized the final person-level analysis weight. All respondents were in the domain for this stage. For the RP model, a domain member was considered an item respondent if all four base variables (CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN) were nonmissing. No respondents received skip codes for the final imputation-

¹⁸⁸ The health insurance method varies slightly from other methods used to impute variables for NSDUH. First, it uses some uncommon constraints and covariates. For example, from the interviewer-administered demographics section, the SERVICE variable, an indicator of service in the United States armed forces, was used both as a covariate and in a likeness constraint related to CHAMPUS. Second, age groups were created to ensure reasonable domain sizes for both modeling and hot-deck steps. Finally, provisional imputation steps in imputation set 1 were not actually hot-deck steps but simple stochastic imputations based on predicted means, and in the final hot-deck step for these variables, different constraints were applied to different age groups.

revised variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. See Appendix D for information on the covariates used in the RP models for health insurance variables.

13.3.1.2 First Prediction Step (CAIDCHIP)

CAIDCHIP was a dichotomous base variable. After imputation, its only values were “yes” and “no.” Therefore, it was modeled using logistic regression, specifically the RLOGIST procedure in SAS®-callable SUDAAN® (RTI International, 2013). The single predicted mean from this model (the probability that the respondent received Medicaid or was covered by a state children’s health insurance program) is used in the final hot-deck step.

13.3.1.3 First Provisional Stochastic Imputation Step (CAIDCHIP)

The first provisional imputation step used for CAIDCHIP was not a hot-deck step, but a stochastic imputation. Each recipient was assigned a “yes” response for the intermediate imputation-revised variable INTCDCDCH with probability equal to the predicted mean from the model. This method was based on the idea of a centered PMN, as described in Singh et al. (2004). As applied to single dichotomous variables like CAIDCHIP, the centered PMN approach reduces to a simple stochastic imputation that requires no donors and no constraints. This approach to imputation for categorical variables is further explored in the *Evaluation of Imputation Methods for the National Survey on Drug Use and Health* report (Center for Behavioral Health Statistics and Quality, 2017).

13.3.1.4 Analogous Prediction and Provisional Hot-Deck Imputation Steps (MEDICARE, CHAMPUS, and PRVHLTIN)

PRD models were fit for MEDICARE, CHAMPUS, and PRVHLTIN in the same manner as for CAIDCHIP, as described above, with a few notable deviations:

- For the MEDICARE variable, a single PRD model was fit for the age group of 18 to 64. This was done because (1) only a small proportion of respondents in these age groups had Medicare, particularly for the age group of 18 to 25, and (2) respondents of working age could have received Medicare only if they were not working because of disability. This was true regardless of whether respondents were aged 18 to 25 or 26 to 64.
- The CHAMPUS PRD models used a covariate that underwent a simplified imputation procedure. For respondents aged 18 or older, models included an indicator of whether the respondent had ever been in the military service, designated by an imputation-revised version of the edited variable SERVICE. The variable SERVICE generally had a very low level of missingness. Although covariates in these models were not supposed to have any missing values, missing values were observed in the SERVICE variable. In those situations, they were randomly imputed as “yes” responses if a generated random number was greater than the mean value of SERVICE across all the other respondents, and they were imputed as “no” responses otherwise.
- PRVHLTIN did not undergo a provisional imputation step because it was the last variable in the set.

13.3.1.5 Final Hot-Deck Step

This section explains the general ideas behind the hot-deck step applied to these variables. Details on the missingness patterns, constraints, and predictive mean vectors for the final hot-deck step of this first stage of imputation are available in Section E.7 in Appendix E.

There are 15 missingness patterns for the first stage of health insurance imputation.¹⁸⁹ No logical constraints were applied, because there were no logical relationships among the variables (or between the variables and any other NSDUH variables). The predictive mean vectors were also made simple because of the absence of logical relationships. Nonetheless, the likeness constraints were fairly sophisticated and sometimes varied across age groups, exploiting auxiliary variables from the demographics and income sections. The SERVICE variable was also used, but it was not an imputation-revised version. The constraint only applied when the item nonrespondent had a nonmissing value.

13.3.2 Any Other Health Insurance Variable, Imputation Set 2

The fifth and final imputation-revised variable, IROTHHLT, was created from the base variable ANYOTHER. This variable indicated whether respondents had any other type of health insurance, even though they reported or were imputed to have none of the four specific types of health insurance, as recorded by IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. The PMN imputation type used for this variable was single RP/single PRD, and there were no noteworthy deviations from this general approach.

For this second stage, the age groups were 12 to 17, 18 to 25, and 26 or older. Three age groups were used instead of four because of the small number of respondents who would have otherwise comprised the domain for the age group of 65 or older.

13.3.2.1 Response Propensity Step

The RP models for imputation set 2 utilized the final person-level analysis weight. The domain for the ANYOTHER variable included respondents who had either a reported or imputed “no” value to all four imputation-revised specific health insurance variables from the first stage (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT). For domain members to be considered item respondents, they had to have complete data for the variable ANYOTHER. See Appendix D for information on the covariates used in the RP models for health insurance variables.

13.3.2.2 Prediction Step

Like all the other health insurance variables that undergo imputation, ANYOTHER was a dichotomous variable. After imputation, its only values were “yes” and “no.” Like the variables in the first stage, it was modeled using logistic regression, as implemented by the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean that was output by the RP step and used in the later hot-deck step was the predicted probability that respondents received

¹⁸⁹ Each of the four variables in the set can be missing or nonmissing. This leads to $2 \times 2 \times 2 \times 2 = 16$ possible combinations, but the combination where none are missing is one that does not require imputation.

other health insurance, given that they were not covered by Medicaid/CHIP, Medicare, CHAMPUS, or private health insurance.

13.3.2.3 Hot-Deck Step

The hot-deck step for the ANYOTHER variable was the simplest one used in NSDUH. There were no logical constraints, the only likeness constraint was the delta constraint, and the predictive mean vector was a scalar. About 5 percent of the records in the domain underwent imputation, and most item nonrespondents were handled on the first attempt to find a donor.

13.3.3 Health Insurance Recodes for Additional Analyses

In addition to the five imputation-revised variables, the overall health insurance variable, IRINSUR4, was created during the imputation process. This variable was created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT. If respondents had a reported or imputed “yes” value for any of these five variables, they were considered to have health insurance. Though IRINSUR4 was technically a recoded variable created from other variables, an imputation indicator (IIINSUR4) was nevertheless created. IIINSUR4 was set to “1” if respondents had a reported “yes” value for any of the five health insurance variables or if respondents had a reported “no” value for all five of them or was set to “3” otherwise. An additional recoded variable, OTHINS, was created to indicate whether respondents were covered by any kind of health insurance other than private health insurance or Medicaid/CHIP. This variable was created by combining IRMEDICR, IRCHMPUS, and IROTHHLT. If respondents had a reported or imputed “yes” value for any of these three variables, they were considered to have other health insurance. No imputation indicator was created for this recoded variable.

14. Editing and Imputation for the NSDUH Pair Variables

14.1 Introduction

Each household selected for the National Survey on Drug Use and Health (NSDUH) has zero, one, or two household members selected for the survey. When two members of the same household (i.e., a pair) are selected, the data can be used to study outcome variables based on the relationship between the respondents (i.e., the pair relationship). The outcome variables can be at either the person level or the pair level. The most common type of analysis is the person-level analysis, where the inferential population is defined by one of the pair members (called the focus pair member). An example of an outcome at the person level is the proportion of children aged 12 to 17 years old who use drugs and whose parents report talking to them about drugs. The focus is on the child in a parent-child pair. By contrast, an example at the pair level is the proportion of parent-child pairs among all possible parent-child pairs where the child is aged 12 to 17 and both the parent and the child smoke cigarettes. This chapter describes the techniques used to edit and impute the pair variables from the household roster (hereafter, “pair variables”) for the 2023 NSDUH.¹⁹⁰ The variables described in this chapter were edited and imputed in three stages:

- Stage One: Pair Relationships,
- Stage Two: Multiplicity Counts, and
- Stage Three: Household-Level Person Counts.

These variables were divided in this way because the variables from earlier stages were needed for the creation of variables in later stages. Stage one comprised the creation and imputation of the variables that identify the pair relationships. The multiplicity and household-level person counts were created and imputed in stages two and three, respectively. Missing values in all three stages were imputed using the predictive mean neighborhood (PMN) imputation procedure, which uses predicted means from models to find donors in a nearest neighbor hot deck. Chapter 3 of this report provides background information about imputation in general (including hot-deck imputation) and details about the PMN methodology in particular.

Though this chapter first presents the editing procedures for each stage and then the imputation procedures for each stage, it is important to note that the actual order of processing was by stage; that is, both editing and imputation were completed for the variables in each stage before moving on to the next stage.

14.1.1 Pair Relationship Variables

The pair relationship variables were derived from the household composition (roster) variables, described in Chapter 11 of this report. Pair relationship variables include the edited and

¹⁹⁰ Edits to the variables for the parenting experiences section, which also involves selection of a pair, are discussed in Section 9.6. The parenting experiences section was administered if (1) an adult and an adolescent pair were selected, and (2) the adult was the parent of the adolescent who also was selected.

imputation-revised pair relationship variables PAIRREL¹⁹¹ and IRPRREL, as well as the imputation indicator IIPRREL, which summarizes how the data in IRPRREL were obtained. PAIRMEM was created to indicate that a respondent was part of a respondent pair. RELMATCH was created to detail how well the pair members matched on their relationships to one another. PRNTIND was created to identify whether the respondent was a parent in a parent-child relationship. Finally, three additional variables were created to aid in pair analyses by giving details about the other pair member: AGEOTHER, SEXOTHER, and PAIRID. AGEOTHER contains the age of the other respondent in the pair; SEXOTHER contains the gender of the other respondent in the pair; and PAIRID contains the questionnaire ID (QUESTID) of the other pair member.

14.1.2 Multiplicity Count Variables

When analyses were conducted at the pair level, the pair domain was completely defined by the pair relationship. For example, to tabulate the number of sibling-sibling pairs where both siblings had used marijuana, it was necessary to know only whether a pair of respondents contained two siblings. “Multiplicity” is the term used to describe the complication that arises in the analysis of pair data at the person level for a given pair domain. For person-level analyses, several pairs in the household could be associated with the same person. The multiplicity count is a count of the number of pairs in the household that can be associated with the person of focus (i.e., the parent or the child of a parent-child pair). As an example, consider a household where a 12-year-old boy and his mother could be selected for the survey and the domain of interest was parent-child pairs where the child was aged 12 to 14. If the boy’s father also lived in the house, then the child-focus multiplicity count would be two, because there were two possible parent-child pairs involving this child. If the boy had a twin brother and a 14-year-old sister, then the parent-focus multiplicity count would be three, because any one of the three children in this age group could have been selected along with the mother.

If multiplicity was not considered, then individuals of focus who were members of more pairs received a weight that was too large (i.e., they were counted too many times in the target population), and individuals of focus who were members of fewer pairs received a weight that was too small. Adjustment for multiplicity was not done a priori because that would lead to the addition of 12 weights, one for each of the 12 possible pair relationships that would require this adjustment (e.g., excluding the spouse-spouse and spouse-spouse with children relationships). Note that the multiplicity problem did not arise if there was only one inclusion possibility (e.g., a single-parent household, if the child was the focus) or if the analysis was a pair-level analysis (e.g., parent-child pair drug behavior).

To illustrate how multiplicities appear in the definitions of parameters and estimates, consider estimating the total number of children who used drugs in the past year, where a parent reported talking to them about drugs. Let $y_{hip}(d)$ be defined as the drug-related behavior outcome for pair p containing the individual i belonging to domain d in household h . Now, for

¹⁹¹ The levels of PAIRREL are provided in [Table 14.3](#).

the population of all individuals who belong to the domain d , the total estimate for drug-related behavior, $\tau_y(d)$, is defined as

$$\tau_y(d) = \sum_{h=1}^H \sum_{i=1}^{N_h(d)} \sum_{p=1}^{M_{hi}(d)} \frac{y_{hip}(d)}{M_{hi}(d)}$$

(i.e., total of averages over pairs (p) associated with the individual i , over all i in domain d and in the household h) (Chromy & Singh, 2001). Here $M_{hi}(d)$ denotes the multiplicity (i.e., the number of pairs associated with the individual i in domain d), $N_h(d)$ can be thought of as the multiplicity count for the household h (i.e., the number of people in the household that are in domain d), and H denotes the total number of households.

Recall the motivating example of estimating $\tau_y(d)$ is equal to the total number of children who used drugs in the past year where a parent reported talking to them about drugs. To estimate this total, the formula components are defined as follows:

d = parent-child pairs with a child focus

$y_{hip}(d)$ = child i used drugs in the past year where a parent from pair p reported talking to the child about drugs in household h (either a 0 or 1 value)

$M_{hi}(d)$ = number of possible parent-child pairs for child i in household h

$N_h(d)$ = number of children in a parent-child pair in household h

Multiplicity count variables were only created for specific relationships of interest. These variables are listed in [Table 14.1](#). With the exception of the spouse-spouse multiplicity variables listed at the bottom of this table, these variables all underwent imputation.

Table 14.1 Edited and Imputation-Revised Multiplicity Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of parents of a child aged 12-14 who is a member of a parent-child pair	MCPCC14	IRMPCC14
Number of children aged 12-14 belonging to a parent who is a member of a parent-child pair	MCPCP14	IRMPCP14
Number of parents of a child aged 15-17 who is a member of a parent-child pair	MCPCC57	IRMPCC57
Number of children aged 15-17 belonging to a parent who is a member of a parent-child pair	MCPCP57	IRMPCP57
Number of parents of a child aged 12-17 who is a member of a parent-child pair	MCPCC17	IRMPCC17

**Table 14.1 Edited and Imputation-Revised Multiplicity Count Variables
(continued)**

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of children aged 12-17 belonging to a parent who is a member of a parent-child pair	MCPCP17	IRMPCP17
Number of parents of a child aged 12-20 who is a member of a parent-child pair	MCPCC20	IRMPCC20
Number of children aged 12-20 belonging to a parent who is a member of a parent-child pair	MCPCP20	IRMPCP20
Number of siblings aged 12-14 for a respondent aged 15-17 who is a member of a sibling-sibling pair	MCS1417	IRMS1417
Number of siblings aged 15-17 for a respondent aged 12-14 who is a member of a sibling-sibling pair	MCS1714	IRMS1714
Number of siblings aged 12-17 for a respondent aged 18-25 who is a member of a sibling-sibling pair	MCS1725	IRMS1725
Number of siblings aged 18-25 for a respondent aged 12-17 who is a member of a sibling-sibling pair	MCS2517	IRMS2517
Pair relationship is spouse-spouse	MCSPSP	N/A
Pair relationship is spouse-spouse with children younger than 18	MCSPSPWC	N/A

N/A = not applicable.

NOTE: For details on multiplicity count variables, see the [NSDUH Pair Manual](#).

14.1.3 Household-Level Person Count Variables

The household-level person count variables represented the number of individuals of focus in the household for a given pair domain, regardless of which pair (or whether a pair) was selected. Household person counts for all domains were calculated for every respondent, even when only one respondent was selected in the household or when a selected pair did not fall in a particular domain. For example, if two parents were in the household with three children aged 12 to 14, the household person count for the parent focus of the parent-child (12-14) domain would be 2 (e.g., there are 2 parents in the household and each could form a parent-child pair), and the household person count for the child focus would be 3 (e.g., each child could be part of a parent-child pair). If the parents indicated that they had a spousal relationship, the household person count for the spouse-spouse domain and the spouse-spouse with children domain would be 1. The rest of the household counts would be zero (even the sibling focus) because they did not fall in the age groups of the domains that were edited and imputed; that is, counts were not created for siblings who both were aged 12 to 14. The household-level person count variables are listed in [Table 14.2](#), and all underwent imputation.

Table 14.2 Edited and Imputation-Revised Household-Level Person Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of children aged 12-14 in the household with at least one parent living with them	HCPCC14	IRHPCC14
Number of parents in the household with at least one child aged 12-14 living with them	HCPCP14	IRHPCP14
Number of children aged 12-17 in the household with at least one parent living with them	HCPCC17	IRHPCC17
Number of parents in the household with at least one child aged 12-17 living with them	HCPCP17	IRHPCP17
Number of children aged 12-20 in the household with at least one parent living with them	HCPCC20	IRHPCC20
Number of parents in the household with at least one child aged 12-20 living with them	HCPCP20	IRHPCP20
Number of household members aged 15-17 with a sibling aged 12-14 living with them	HCS1417	IRHS1417
Number of household members aged 18-25 with a sibling aged 12-17 living with them	HCS1725	IRHS1725
Number of spouse-spouse pairs	HCSPSP	IRHCSPSP
Number of spouse-spouse pairs with children younger than 18	HCSPSPWC	IRHCSPWC

14.2 Stage One Editing: Pair Relationships

14.2.1 Editing the Household Roster of Each Pair Member

Prior to identifying the relationships between selected pair members, a key step was editing the questionnaire household rosters for each pair member. This involved identifying situations where the relationship listed in the roster for a particular roster member was not possible given the roster member's age and relationship to the respondent. Usually, the relationships could not be determined and the relationship code was set to bad. In general, no effort was made to try to match the values of roster-derived household composition variables between pair members, because different members of the same household could have completed the survey at different times. However, information from other pair members was sometimes used to change a relationship code or to change the age of a roster member to match the age of the other respondent.

14.2.2 Creating the Pair Relationship Variable (PAIRREL)

The process of identifying the pair relationships involved two steps: (1) match the household rosters of the pair members, and (2) determine the pair relationships using the relationship codes and ages of the matched rosters. The first step is described in Section 14.2.2.1 and Section H.1 of Appendix H, and the second step is described in Section 14.2.2.2 and Section H.2

of Appendix H. Any relationships that could not be determined by this process were imputed as described in Section 14.5.2.

No restrictions were placed on the types of pairs that could be selected for inclusion in the NSDUH sample. However, the variable PAIRREL includes a limited number of relationships. The identification of the particular relationships between a given pair was originally limited by the relationship codes that were available: parent, child, grandparent, grandchild, sibling, spouse, unmarried partner, roommate, parent-in-law, child-in-law, boarder, other relative, and other nonrelative.

The following pair relationships were considered to be of higher analytic interest and were used to inform the definitions of the levels of PAIRREL:

- parent-child, child aged 12 to 14;
- parent-child, child aged 12 to 17;
- parent-child, child aged 15 to 17;
- parent-child, child aged 12 to 20;
- sibling-sibling, younger sibling aged 12 to 14, older sibling aged 15 to 17;
- sibling-sibling, younger sibling aged 12 to 17, older sibling aged 18 to 25;
- spouse-spouse (includes partner-partner), with children younger than 18;¹⁹² and
- spouse-spouse (includes partner-partner), with or without children.

The levels in PAIRREL, summarized in [Table 14.3](#), do not correspond exactly with the pairs of interest given above, but the relevant pair relationships can be derived from the value of PAIRREL. For example, a value of PAIRREL = 3 indicates that, among the pair relationships given above, the pair relationship was a parent-child pair with a child aged 18 to 20.

Table 14.3 Levels of the Variable PAIRREL

Value of PAIRREL	Interpretation	Domain of Interest
1	Respondent is part of a parent-child (12-14) pair	Yes
2	Respondent is part of a parent-child (15-17) pair	Yes
3	Respondent is part of a parent-child (18-20) pair	Yes, indirectly
4	Respondent is part of a parent-child (21+) pair	No
5	Respondent is part of a sibling (12-14)-sibling (15-17) pair	Yes
6	Respondent is part of a sibling (12-17)-sibling (18-25) pair	Yes
7	Respondent is part of another sibling-sibling pair	No

¹⁹² The spouse-spouse pair relationship included respondents who were legally married, as well as respondents who lived together as though married (partners). Although the questionnaire distinguished between “spouses” and “partners,” the pair relationship variable being described here did not distinguish between the two. In rare instances, a spouse-spouse pair included one pair member who identified the second pair member as a spouse, whereas the second pair member identified the first as a partner.

Table 14.3 Levels of the Variable PAIRREL (continued)

Value of PAIRREL	Interpretation	Domain of Interest
8	Respondent is part of a spouse-spouse ¹ pair, with children in the household younger than 18	Yes
9	Respondent is part of a spouse-spouse pair, with no children in the household younger than 18	Yes
10	Respondent is part of a spouse-spouse pair, but it is unclear whether children younger than 18 in the household belong to the pair	Yes
11	Respondent is part of a grandparent-grandchild pair	No
12	Respondent is part of another clearly identifiable pair	No
13	Respondent is part of a pair that is not clearly identifiable, but it is clear from the relationship codes that it is not within codes 1 through 11	No
14	Respondent is part of a pair that is not clearly identifiable, and it could be any pair relationship	Maybe
15	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 1 or 12	Maybe
16	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 2 or 12	Maybe
17	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 3 or 12	Maybe
18	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 4 or 12	No
19	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 5 or 12	Maybe
20	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 6 or 12	Maybe
21	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 7 or 12	No
22	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8 or 12	Maybe
23	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 9 or 12	Maybe
24	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8, 9, or 12	Maybe
25	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 11 or 12	No
99	Respondent is not a member of a pair	No

¹ The pair relationship labeled “spouse-spouse” includes partner-partner pair relationships.

14.2.2.1 Matching the Household Rosters

For the purpose of discussing how to match the household rosters in this report, the pair members were identified as pair member “A” and pair member “B.” For the household roster of pair member A, it was necessary to determine which household member listed in pair member A’s roster corresponded to the other selected pair member. The same had to be done for pair member B. This was accomplished using the age and gender of the pair members, in addition to the variable MBRSEL, which was used to identify the roster member corresponding to the other selected pair member.

In a perfect setting, the questionnaire age and gender of pair member B (AGE and IRSEX, respectively) would correspond exactly to the age and gender entered for one of the members of pair member A’s household roster (RAGE and RSEX). Moreover, the value of MBRSEL for this matched roster member would be 1, and the value of MBRSEL for all other roster members would be zero or missing. In this ideal setting, matches with exactly one MBRSEL = 1 correctly identifying the other pair member would be found with both pair member A’s and B’s roster. This did not always occur, and some effort was required to determine the roster member most likely to correspond to the other selected pair member.

The quality of the match between pair members varied depending upon the quality of the roster entries and the time between survey completions. Several conditional tests (e.g., if-then-else), called priority conditions (because of the hierarchical nature of the conditions), were used to determine the validity of a pair match. These conditions are provided in Appendix H. In general, the conditions matched IRSEX and AGE for the one pair member against the gender and age of the roster members in the other pair member’s roster, using MBRSEL to help identify the appropriate roster member. For a match to be considered valid, it was necessary that at least one of the two pair members have a match.

When a single roster member had to be selected among duplicates and the duplicates had the same relationship code, the other pair member was selected at random. Sometimes because of the poor quality of the rosters of the pair members, it was not possible to locate the household member listed in pair member A’s roster that corresponded to pair member B, and vice versa. The determination of the pair relationships for these situations was left to imputation. Even when a pair of roster members was successfully identified, it was not always possible to successfully determine the pair relationship, as described in the next section.

14.2.2.2 Determining the Pair Relationship Using the Relationship Codes of the Matched Rosters

The identification of the pair resulted in two observations per household, each with a relationship code corresponding to the other selected pair member. The relationship codes for these two observations had to be matched to determine the pair relationship. For example, suppose a 15-year-old and a 38-year-old were selected. If the 38-year-old was identified as the parent on the 15-year-old’s roster, and the 15-year-old was identified as the child of the 38-year-old on the 38-year-old’s roster, then the pair relationship would be identified as PAIRREL = 2, according to the levels of PAIRREL provided in [Table 14.3](#).

As with the procedure used to match the household rosters, a series of conditions was used to identify the relationship between pair members. These priority conditions used ages and relationship codes to identify the pair relationships. In a perfect setting, the relationship codes would be nonmissing and in agreement between the pair members, as in the example given in the previous paragraph. In some instances, however, either the relationship codes were missing or they did not agree across the pair members. The priority conditions offered a method for editing the relationship codes when this occurred.

Below are the strategies used to identify a pair relationship in an imperfect setting:

1. If a relationship code was missing on one side of the pair but not on the other, then the pair relationship was assumed to be identified by the nonmissing relationship code. The exception to this rule occurred if the identified relationship was parent-child with a child younger than 18, the "parent" was less than 10 years older than the child, and the "parent" answered the parenting experiences question (FIPE3) by saying that the other respondent was not the parent's child. In this situation, the nonmissing relationship code was considered spurious, and the relationship was left missing.
2. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but the relationship codes on both sides indicated that the unknown pair relationship was not a relationship of interest, then the pair relationship was identified as such and no imputation was required. For example, if pair member "A" identified pair member "B" as a "boarder," but pair member "B" identified pair member "A" as "other relative," then the relationship was not a relationship of interest and the variable PAIRREL would have been assigned the value "13" ([Table 14.3](#)).
3. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but a parent-child relationship was possible given the relationship code in one of the pair member's rosters, then the FIPE3 variables were used to assist in the determination of a pair relationship. For example, consider a pair member who was a stepparent and refers to the stepchild as "child," but the child refers to the stepparent as "other nonrelative." Membership in a parent-child relationship where the child was younger than 18 was indicated if the stepparent answered FIPE3 affirmatively, thereby proceeding to the parenting experiences section. However, if the stepparent answered FIPE3 negatively, then the stepparent was not considered the parent. A third scenario arose if the FIPE3 answer was not given. In this situation, a parent-child relationship was assumed if the stepparent was legally married and the child identified the spouse of the other pair member as "parent."

The quality of the match for PAIRREL levels 1 through 25 is indicated by the variable RELMATCH, the levels of which are summarized in [Table 14.4](#). In general, imputation was required for values of RELMATCH of 0 or 4, or if PAIRREL = 10. PAIRREL = 10 was a special circumstance because the relationship between the pair members was clear, but the presence of children younger than 18 belonging to the pair was not clear. For this value of PAIRREL, the value of RELMATCH = 1 or 2, but imputation was still required. The number that was matched or not matched, as indicated by the RELMATCH variable (or PAIRREL = 10), for the 2023 survey

is provided in [Table 14.5](#). The amount of imputation required was dependent upon the quality of the rosters. The attributes of the roster are described in Section 11.1.

Table 14.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH

Value of RELMATCH	Values of PAIRREL	Interpretation
0	14	FAILURE: The relationship was not identifiable and could have been a relationship of interest.
1	1-9, 11-13	SUCCESS: The relationship was clearly identifiable using information from both pair members or was unmistakably not a relationship of interest.
1	10	FAILURE: A spouse-spouse ¹ relationship was definitively established using information from both pair members, but it was unclear whether the pair had children in the household.
1.5	8	SUCCESS: A spouse-spouse relationship was definitively established using information from both pair members, and children younger than 18 were in both rosters. Relationship codes on one side of the pair indicated children belonged to the pair, and on the other side of the pair, the relationship codes corresponding to the children were missing.
2	1-9, 11-13	SUCCESS: The relationship was clearly identifiable using information from one pair member, while the relationship code from the other pair member was missing.
2	10	FAILURE: A spouse-spouse relationship was definitively established using information from one pair member, while the relationship code from the other pair member was missing. It was unclear whether the pair had children in the household.
3	1-8, 12, 13	SUCCESS: Relationship information was conflicting between the pair members, but conclusions were drawn anyway for some parent-child pairs, some sibling-sibling pairs, and some spouse-spouse pairs using either information outside the household roster or logical reasoning. ²
4	15-25	FAILURE: Relationship information was not identifiable. Information was in conflict between the pair members, where one pair member indicated relationship of interest and the other did not. However, ages supported a relationship of interest (may be used to limit imputation).

¹ The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

² In the situation of potential parent-child pairs, further evidence that a parent-child relationship was involved or not involved was obtained by looking at the FIPE3 variable to see whether a stepparent had a spouse who corresponded to a child's parent or to see the ages of the respondents. For spouse-spouse relationships, two situations occurred. When the respondents were not legally married, the children of one pair member were considered the children of the pair in the household, even though they were not identified as such by the other pair member. When only one pair member referred to the other as a "married" or "unmarried partner," and if both had the same children, they were considered "spouse-spouse-with-children." The other pair member was usually referred to as a "roommate" or "other nonrelative."

Table 14.5 Frequencies of the Levels of the Variable RELMATCH: 2023

RELMATCH	Frequency (Percent)	Requires Imputation
0	71 (0.53)	Yes
1 (PAIRREL ≠ 10)	12,802 (95.84)	No
1 (PAIRREL = 10) ¹	46 (0.34)	Yes
1.5	4 (0.03)	No
2 (PAIRREL ≠ 10)	233 (1.74)	No
2 (PAIRREL = 10)	0 (0.00)	Yes
3	51 (0.38)	No
4 ²	151 (1.13)	Yes

¹ When PAIRREL = 10, the respondent is part of a spouse-spouse pair, but it is unclear whether children younger than 18 in the household belong to the pair. Imputation is required to determine whether the spouse-spouse pair has children.

² When RELMATCH = 4, imputation is required to determine the pair relationship because only one pair member indicated a relationship of interest and the ages supported a relationship of interest.

14.3 Stage Two Editing: Multiplicity Counts

As stated earlier, multiplicities were required to account for analyses that were performed at the person level, in addition to the pair weights that were calculated at the pair level. Because the multiplicities were relevant only at the person level, the definition of multiplicity required the identification of the focus member of the pair. Because the creation of the multiplicity factors required complicated programming logic, multiplicities were not created for all possible pair relationships. Only certain pair relationships considered to be of higher analytic interest were included in the creation of multiplicities. Using the pair relationships determined in Section 14.2, the following domains were considered:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 15 to 17), parent focus;
4. parent-child (child 15 to 17), child focus;
5. parent-child (child 12 to 17), parent focus;
6. parent-child (child 12 to 17), child focus;
7. parent-child (child 12 to 20), parent focus;
8. parent-child (child 12 to 20), child focus;
9. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
10. sibling (12 to 14)-sibling (15 to 17), sibling (12 to 14) focus;
11. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
12. sibling (12 to 17)-sibling (18 to 25), sibling (12 to 17) focus;
13. spouse-spouse (includes partner-partner) with children younger than 18; and
14. spouse-spouse (includes partner-partner).

Determining the multiplicity entailed finding the number of roster pairs in the domain of interest that contained the focus member in the pair. In broad terms, the process of determining the

multiplicity count involved two steps: (1) calculate the multiplicity count for each pair member, and (2) use the screener, quality of roster, and other means to determine the appropriate count if pair members' counts did not match. The first step is described in Section 14.3.1, and the second step is described in Section 14.3.2 and Appendix I. Multiplicities that could not be determined through these steps were left to imputation as described in Section 14.5.3.

Because the pair weights reflected selection done at the time of screening, the multiplicity count should reflect the household makeup at that time. However, this was not always possible, because for some households the screener roster was not as complete as the questionnaire roster, and recorded relationships in the screener roster were relative to the head of the household for in-person interviews and to the screener respondent for web interviews rather than to each pair member. No account was made for situations where a change in the household makeup occurred between the screening date and the survey completion date. Due to the passage of time only, the observed change in household makeup could have occurred because of an intervening birthday or because a roster member left or entered the household after screening. Because an adjustment to the multiplicity counts would have been extremely complicated to implement for the small number of occurrences, no such adjustment was made. Nevertheless, when possible, the screener was used to resolve disagreements between pair members on the value of the multiplicity count.

14.3.1 Determining the Multiplicity Count for Each Pair Member

The multiplicity counts for each pair member consisted of a direct count and an indirect count. The direct count was obtained by looking at the focus pair member. It was a count of the roster members who could have been selected, where the same pair domain would have resulted. The indirect count was obtained by looking at the pair member who was not the focus. It was a count of the pair member, plus other roster members who, by virtue of their relationship code, would have had the same pair relationship had they been selected.

A summary of the ways of determining the direct count and indirect count for each pair domain are provided in [Table 14.6](#). For these domains, neither the direct nor the indirect count could be zero, because the pair member who was not the focus had to be part of the count. For spouse-spouse counts, no work was necessary to determine multiplicity counts. If a respondent was in a spouse-spouse pair, the multiplicity count was necessarily one in almost all situations, because only one spouse-spouse pair could have been selected that included that pair member. If the true multiplicity count exceeded one, then the multiplicity count was set to one.¹⁹³ Note that other spouse-spouse pairs in the household (e.g., one spouse's parents) would have been of interest in the household counts discussed in subsequent sections.

¹⁹³ In rare circumstances, it was possible for a respondent to have two or more spouses who lived in the same household. Determining the appropriate multiplicity count in these situations required knowledge of which spouse was the focus, which would be arbitrary. Because having multiple spouses was an extremely rare occurrence, and because of the complexity of determining the appropriate multiplicity count, these situations were not explicitly addressed during data processing.

Table 14.6 Multiplicity Counts for Each Pair Member

Pair Relationship	Focus Member	Direct Count	Indirect Count
Parent-Child	Child	From child: number of parents	From parent: self + spouse/partner
Parent-Child	Parent	From parent: number of children in appropriate age range	From child: self + number of siblings in the appropriate age range
Sibling-Sibling	Older Sibling	From older sibling: number of siblings in younger age range	From younger sibling: self + number of siblings in younger age range
Sibling-Sibling	Younger Sibling	From younger sibling: number of siblings in older age range	From older sibling: self + number of siblings in older age range

14.3.2 Determining the Final Multiplicity Count

Once the counts were determined for each pair member, it was necessary to resolve any differences between these counts across pair members. Most of the time, the direct and indirect counts agreed, with no bad relationship codes for either pair member, resulting in a straightforward determination of the final multiplicity count. If one pair member had bad relationship codes or had a count of zero, the final multiplicity count came from the pair member with good data.¹⁹⁴ Sometimes both pair members had bad relationship codes, which meant that the final multiplicity was left to imputation. Among the remaining situations, some could be reconciled and some could not. When reconciliation was possible, many of the disagreements between the pair members were resolved by going to the screener. The method used to reconcile differing counts depended upon the domain. For the parent-child domains, for example, in addition to the screener, the parent/legal guardian variable, FIPE3, was used to help reconcile differences. Detailed rules for reconciling differences between pair members are provided in Appendix I.

If reconciliation between the counts from the two pair members in the household and the screener was not possible, upper and lower bounds within which the imputed value had to reside were determined from the counts for each pair member and the counts for the screener. The amount of imputation required for the multiplicity counts for the 2023 survey is shown in [Table 14.7](#). The greatest degree of uncertainty came with the determination of the number of parents in the child-focus parent-child domains. This occurred because, even though the parent-child pair relationship had been established, often it was unclear whether there was a second parent in the household.

Other domains had very little uncertainty. The counts of the number of children in the parent-focus parent-child domains, for example, were almost always definitively determined.

¹⁹⁴ There were some exceptions to this rule. If the bad relationship codes were only within the relevant age ranges, then the count from the good side was used only if the age ranges in the good side matched the screener.

Table 14.7 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2023

Pair Domain	Multiplicity	Number Missing
Parent-Child (12-14), Child Focus	Number of parents	98
Parent-Child (12-14), Parent Focus	Number of children	73
Parent-Child (15-17), Child Focus	Number of parents	105
Parent-Child (15-17), Parent Focus	Number of children	78
Parent-Child (12-17), Child Focus	Number of parents	132
Parent-Child (12-17), Parent Focus	Number of children	87
Parent-Child (12-20), Child Focus	Number of parents	148
Parent-Child (12-20), Parent Focus	Number of children	102
Sibling (12-14)-Sibling (15-17), Older Sibling Focus	Number of younger siblings	72
Sibling (12-14)-Sibling (15-17), Younger Sibling Focus	Number of older siblings	72
Sibling (12-17)-Sibling (18-25), Older Sibling Focus	Number of younger siblings	75
Sibling (12-17)-Sibling (18-25), Younger Sibling Focus	Number of older siblings	77

14.4 Stage Three Editing: Household-Level Person Counts

To improve the quality of the estimates from the pair data through poststratification of the appropriate weights,¹⁹⁵ it was necessary to identify the household-level person counts for each domain. This entailed finding the number of individuals in the household who belonged to a particular domain, given that one member of a domain was selected as the focus. These counts were more difficult to derive than the multiplicity counts because all households were considered. Within each household, counts for any of the domains of interest were derived, regardless of whether the pair belonged to that domain or even whether a pair was selected at all. The counts were derived for 10 of the 14 pair domains described in Section 14.3. For the parent-child counts where the child was aged 15 to 17, calculating the household counts was unnecessary.¹⁹⁶ The other two remaining sibling-sibling domains were added after the procedures were first developed, and there was insufficient time to develop the household counts for those domains. The domains where these counts were created are listed below:

1. parent -child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;

¹⁹⁵ See the *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 12: Questionnaire Dwelling Unit-Level and Person Pair-Level Sampling Weight Calibration* report (Center for Behavioral Health Statistics and Quality, 2025c) for details.

¹⁹⁶ Because household counts were defined for everybody, it was possible to derive these counts using the counts for the parent-child domains where the child was aged 12 to 14 and where the child was aged 12 to 17. However, the multiplicity counts for the parent-child (15 to 17) domain had to be calculated and could not have been derived in as straightforward a way, because the multiplicity counts were only defined if the pair relationship corresponded to the pair domain of interest.

3. parent-child (child 12 to 17), parent focus;
4. parent-child (child 12 to 17), child focus;
5. parent-child (child 12 to 20), parent focus;
6. parent-child (child 12 to 20), child focus;
7. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
8. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
9. spouse-spouse (includes partner-partner) with children younger than 18; and
10. spouse-spouse (includes partner-partner).

Determining the household-level person counts was a two-step process: (1) calculate the household count for each respondent, whether a member of a pair or a single respondent; and (2) use the screener, quality of roster, and other means to determine the appropriate final count either by attempting to reconcile differing counts between pair members or by attempting to determine the appropriate count when information from only one roster was available. For households where only one respondent was selected, the matching step (step 2) was unnecessary. The first step is described in Section 14.4.1, and the second step is described in Section 14.4.2 and Appendix J. Household counts that could not be determined by this process were left up to imputation as described in Section 14.5.4.

Because the pair weights reflected selection at the time of screening, the household-level person counts should have reflected the household makeup at that time. As with the multiplicity counts, this was not entirely possible. No account was made for situations where a change in the household makeup occurred between the screening time and the time of survey completion by each pair member. An explanation for why this was not possible for the multiplicity counts is described in Section 14.3. Moreover, as stated in that section, to implement such an adjustment would have been extremely complicated for the household-level person counts. Nevertheless, when there were disagreements between pair members on the value of the household-level person count, the screener was used to resolve those disagreements.

14.4.1 Determining the Household-Level Person Count for Each Respondent

14.4.1.1 Parent-Child Domains

In any household, the household-level person counts for parent-child domains were nonzero if at least one parent was present in the household with children within the relevant age range. In this instance, the child-focus count was the number of children in the household within that age range that belonged to any parent in the household, and the parent-focus count was the number of parents. If more than one “family unit” (mother and/or father with children) lived within the household, the child-focus counts would have counted children from more than one set of parents, and the parent-focus counts would have counted two or more parents, at least one for each set of children.

One situation where this could occur was when three generations lived within the same household, with children in both the youngest and the second generations within the relevant age range. For example, a 12-year-old child could have parents, grandparents, and a 19-year-old aunt or uncle in the household. In this instance, the parent-child count of the number of

children aged 12 to 20 would have included both the 12-year-old child and the 19-year-old aunt. The count of the number of parents would have included both the parents and the grandparents (and possibly have exceeded 2).

To obtain the household-level person counts for the parent-child domains outlined above, household rosters were examined to determine a lower bound based on clear parent-child relationships (e.g., the respondent is 15 years old and lists a parent in the household). Then, an upper bound was created by adding in relationships that were not clearly defined as parent-child. These relationships generally occurred outside of the respondent's nuclear family. For example, a respondent may have listed an aunt and a cousin. This could be a parent-child relationship, but that is not definitive. [Table 14.8](#) outlines this process for the 12-14 age group. The process is the same for the other age groups. It is important to note that conditions 1 through 5 were applied in order, and only one of them could apply to a household. In contrast, conditions 6 through 14 were additive, and more than one could apply to the same household. Therefore, care was taken to be sure the same household members were not counted multiple times.

After the initial creation, the bounds for the 12-17 age group were adjusted based on the bounds for the 12-14 age group. The household count of children aged 12 to 17 is necessarily greater than or equal to the count of children aged 12 to 14. Similarly, the bounds for the 12-20 age group were adjusted based on the bounds for the 12-17 age group. In households with only one family unit, the household count was set to the corresponding multiplicity count, where available. Finally, if the bounds were equal, or if the upper bound was less than the lower bound, then the initial household counts were set to the lower bound. If the bounds were not equal, then they were either (1) used in the reconciliation process across the pairs, as described in Section 14.4.2, or (2) used as bounds in imputation for nonpair households.

Table 14.8 Obtaining the Bounds for the Parent-Child, Child (12-14) Domain

Condition Number	Description	Lower Bound		Upper Bound		Number of Family Units in Household
		Parent Focus	Child Focus	Parent Focus	Child Focus	
1	No household members aged 12-14	0	0	Lower bound	Lower bound	0
2	Respondent has parents and both children and siblings aged 12-14	Self + spouse + parents	Children aged 12-14 + siblings aged 12-14	Lower bound	Lower bound	2
3	Respondent has parents and is aged 12-14	Parents	Self + siblings aged 12-14	Lower bound	Lower bound	1
4	Respondent is 14+ but has parents and siblings aged 12-14	Parents	Siblings aged 12-14	Lower bound	Lower bound	1
5	Respondent has children aged 12-14	Self + spouse	Children aged 12-14	Lower bound	Lower bound	1

Table 14.8 Obtaining the Bounds for the Parent-Child, Child (12-14) Domain (continued)

Condition Number	Description	Lower Bound		Upper Bound		Number of Family Units in Household	
		Parent Focus	Child Focus	Parent Focus	Child Focus		
6	Respondent does not list children but has a spouse or partner with a possible child aged 12-14	No change	No change	Add self + spouse	Add other relatives + nonrelatives aged 12-14	+1	
7	Respondent has a spouse or partner aged 12-14	Respondent lists parents-in-law	Add parents-in-law	Add spouse	Add parents-in-law	Add spouse + siblings-in-law aged 12-14	+1
	Respondent lists other or nonrelatives who could be spouse's parents	No change	No change	Add other or nonrelatives aged 26+	Add spouse + other or nonrelatives aged 12-14		
8	Respondent does not have a spouse but has parents-in-law with possible children	No change	No change	Add parents-in-law	Add siblings-in-law + other or nonrelatives aged 12-14	+1	
9	Respondent has grandparents and aunts or uncles aged 12-14	No change	No change	Add grandparents	Add aunts and uncles aged 12-14	+1	
10	Respondent has a sibling or sibling-in-law and a niece or nephew aged 12-14	No change	No change	Add siblings and siblings-in-law aged 26+	Add nieces and nephews aged 12-14	+1	
11	Respondent has children or children-in-law and grandchildren aged 12-14	No change	No change	Add children and children-in-law aged 26+	Add grand-children aged 12-14	+1	
12	Respondent has an aunt or uncle and cousins aged 12-14	No change	No change	Add aunts and uncles aged 26+	Add cousins aged 12-14	+1	
13	Respondent has other relatives who could have parent-child relationships based on age	No change	No change	Add other relatives aged 26+	Add other relatives aged 12-14	+1	
14	Respondent has nonrelative who could have parent-child relationships based on age	No change	No change	Add nonrelatives aged 26+	Add nonrelatives aged 12-14	+1	

14.4.1.2 Sibling-Sibling Domains

When the respondent was a member of a sibling-sibling pair of the corresponding age group, and the household contained only one family unit (e.g., one set of parents with children), the sibling-sibling household count was set equal to the corresponding sibling-sibling multiplicity count. Otherwise, the process for obtaining household counts for sibling-sibling domains was

similar to the one for parent-child domains. First, lower bounds were created based on known sibling relationships, then upper bounds were created based on possible sibling relationships. [Table 14.9](#) outlines this process for HCS1417, the count of household members aged 15 to 17 with a sibling aged 12 to 14. The same process was used to determine the bounds for HCS1725. Just as in the creation of bounds for parent-child relationships, conditions 1 through 3 were mutually exclusive, whereas conditions 4 through 11 were additive. When the bounds were not equal, the household count was left missing and the bounds were used when reconciling across pairs or, for nonpair households, for imputation.

Table 14.9 Obtaining the Bounds for the Sibling-Sibling (12-14/15-17) Domain

Condition Number	Description		Lower Bound	Upper Bound
1	Respondent is aged 12-14		Siblings aged 15-17	Lower bound
2	Respondent is aged 15-17 and has siblings aged 12-14		Self + siblings aged 15-17	Lower bound
3	Respondent is 17+ but has siblings in both age groups		Siblings aged 15-17	Lower bound
4	Respondent has children in both age groups		Add children aged 15-17	Add children aged 15-17
5	Respondent has grandchildren in both age groups		No change	Add grandchildren aged 15-17
6	Respondent has a spouse or siblings-in-law in both age groups	Respondent has a spouse aged 12-14 and siblings-in-law aged 15-17	No change	Add siblings-in-law aged 15-17
		Respondent has a spouse aged 15-17 and siblings-in-law aged 12-14	No change	Add spouse + siblings-in-law aged 15-17
		Respondent has siblings-in-law aged 12-14 and 15-17	No change	Add siblings-in-law aged 15-17
7	Respondent has aunts or uncles in both age groups		No change	Add aunts or uncles aged 15-17
8	Respondent has nieces or nephews in both age groups		No change	Add nieces or nephews aged 15-17
9	Respondent has cousins in both age groups		No change	Add cousins aged 15-17
10	Respondent has other relatives in both age groups		No change	Add other relatives aged 15-17
11	Respondent has nonrelatives in both age groups		No change	Add nonrelatives aged 15-17

14.4.1.3 Spouse-Spouse Domains

What is referred to as a “spouse-spouse domain” was derived from spouse-spouse and partner-partner pair relationships. [Table 14.10](#) outlines the process for determining the bounds for spouse-spouse counts. Multiple conditions could have applied to the same household.

Table 14.10 Obtaining the Bounds for the Spouse-Spouse Domain

Condition Number	Description	Lower Bound	Upper Bound
1	Respondent lists a spouse	+1	Lower bound
2	Respondent lists 2 or more parents	Add number of parents divided by 2 and truncated	Lower bound
3	Respondent lists 2 parents-in-law	+1	Lower bound
4	Respondent lists male-female grandparent pairs	Add the minimum of the number of female and male grandparents	Lower bound
5	Respondent has children aged 15+ and children-in-law	No change	Add the minimum of the number of children aged 15+ and the number of children-in-law
6	Respondent lists only one parent and one nonrelative of the opposite sex	No change	+1
7	Respondent lists a sibling and siblings-in-law of similar ages	No change	Add the minimum of the number of siblings and the number of siblings-in-law
8	Respondent lists male-female pairs of other or nonrelatives	No change	Add the minimum of the number of male and the number of female other or nonrelatives

If the bounds did not match, the count of spouses was set to missing and the bounds were used either to reconcile with the other pair member or as bounds for imputation. The count of spouse-spouse pairs with children younger than 18 was largely based on the overall spouse-spouse count. [Table 14.11](#) lists the conditions required for incrementing the number of spouse-spouse pairs with children younger than 18. Some of these conditions were applied to the same household.

Table 14.11 Obtaining the Bounds for the Spouse-Spouse with Children Younger Than 18 Domain

Condition Number	Description	Lower Bound	Upper Bound
1	Respondent is a member of a spouse-spouse pair with children	+1	Lower bound
2	Respondent is not a member of a spouse-spouse pair but lists a spouse and children	+1	Lower bound
3	Respondent lists 2 parents and is younger than 18 or has siblings younger than 18	+1	Lower bound
4	Respondent lists children aged 15+, children-in-law, and grandchildren younger than 18	No change	Add the minimum of the number of children aged 15+ and the number of children-in-law
5	Respondent lists only one parent and one nonrelative of the opposite sex and is younger than 18 or has siblings younger than 18	No change	+1

Once these initial bounds were created, the upper bound could be increased based on the number of children yet to be counted. The spouse-spouse with children count was bounded by both the total number of spouse-spouse pairs and the number of children in the household. For example, if there were two spouse-spouse pairs in the household, but only one child younger than 18, then there could only be one spouse-spouse pair with children. If the final bounds were equal, then the count was set to the lower bound. Otherwise, the bounds were used in the reconciliation across pairs or as bounds for imputation.

14.4.2 Determining the Final Household-Level Person Count

Most of the time, the household structure was simple, and the creation of the final household-level person counts was straightforward. The most difficult situations involved multi-family households. If a pair was selected, and the pair members were parts of different family units, then the resulting sets of counts or bounds, although not usually equal, would generally be consistent. Consider the following example: Respondent A is a 15-year-old male who lists two parents, a 25-year-old sister, a 30-year-old brother-in-law, and a 12-year-old niece. The parent-child, parent focus count for the 12-20 age group would not be definitive based on this one roster, but would have a lower bound of 2 and an upper bound of 4. If the second respondent was the 25-year-old sister, and she also listed two parents, her 15-year-old brother, a 30-year-old spouse, and a 12-year-old child, then the parent-child, parent focus count for the 12-20 age group based on her roster would be 4, which is within the bounds based on the brother's roster. Sometimes comparing to the other pair member's roster did not completely define the household counts. In the example above, if the second respondent was respondent A's mother, who listed her husband, 30-year-old son-in-law, 25-year-old daughter, 15-year-old son, and 12-year-old granddaughter, the two rosters would not completely define the count of parent-child pairs. In other situations, where the pair members gave conflicting information, counts that

were definitive on one side's roster became bounds for the final count. More detailed rules for reconciling differences between pair members are described in Appendix J.

If the attempt to reconcile differences in the household-level person counts between pair members was unsuccessful, upper and lower bounds within which the imputed value must reside were determined from the counts for each pair member and the counts for the screener. Reconciliation was not possible among nonpair households, and any missing household counts were left for imputation.

14.5 Imputation for the Pair Variables

Imputation was required for variables in all three stages, and the imputation models for the pair variables required the inclusion of covariates at the pair level. The creation of these covariates is described in Section 14.5.1. The imputation process for stages one, two, and three are described in Sections 14.5.2 through 14.5.4. For details on the covariates used in the response propensity (RP) and prediction (PRD) models, see Appendix D.

14.5.1 Creation of Covariates for Imputing Pair-Level Variables

Because imputation was performed at the household level rather than at the respondent level, it was necessary to include covariates defined at the household level. Segment-level covariates were used for this purpose because they were automatically defined at the household level, using external information that was constant regardless of when the survey was completed. In addition to these segment-level covariates, information from the questionnaire would also have been useful as modeling variables. The logical choices for questionnaire-derived variables include the household composition variables IRHHSIZE (household size), IRKID17 (number in household younger than 18), IRHH65 (number in household aged 65 or older), and IRFAMSKP (presence of other family members in household indicator).

Because pair members could have responded at different times, these variables were not necessarily consistent across pair members. Therefore, new count variables were needed that were consistent within a household. These variables were created in a two-step process: (1) create the count variables for each pair member, and (2) attempt to reconcile disagreeing values between pair members. The following sections describe these two steps in the creation of household size, household composition age count variables, and household composition age count variables "for males only," each of which were consistent across pair members. Note that household composition age count variables were not created for females because, for a given age range, the number of females could be obtained by subtracting the number of males from the total number within that age range. These variables also had to be created for respondents who were not part of a pair, for the purposes of creating and imputing the household-consistent person counts of various domains. The reconciliation process was not necessary for nonpair households, and the counts were obtained directly from the respondents' rosters. If a respondent in a nonpair household had missing or bad data, the counts were reconciled against the screener roster.

14.5.1.1 Household Size

The variable created to represent a household size that was consistent across selected pair members was called HHSIZE. This variable was based on the imputed household size IRHHSIZE, the screener household size, and modified counts of the roster members on each side called TGOODAGA and TGOODAGB. Each modified count is equal to the total number of occurrences in the roster with valid ages, incorporating the minimum possible counts within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. HHSIZE was set based on the conditions outlined in [Table 14.12](#). Note that the descriptions in this table refer to sides A and B of the pair. These same conditions applied no matter which pair member was defined as side A.

Table 14.12 Conditions Used to Reconcile HHSIZE across Pairs

Condition Number	Description	Decision	Number of Occurrences in 2023
1	Values for IRHHSIZE agree for sides A and B	Use value from side A	12,071
2	Side A has IRHHSIZE ≤ 1 ; side B has IRHHSIZE > 1 ¹	Use value from side B	0
3	Side A has IRHHSIZE $< TGOODAGA$; side B has IRHHSIZE $\geq TGOODAGB$	Use value from side B	81
4	Both sides have IRHHSIZE $< TGOODAGA$	Use maximum of TGOODAGEA, TGOODAGEB, and the screener count	2
5	Screener count matches the sum of the maximum of the counts for age groups 0-11, 12-14, and 15+	Use screener count	508
6	Side A is closer to and larger than the screener count	Use value from side A	595
7	Side A is closer to the screener count	Use value from side A	47
8	Neither side is closer to the screener count	Use the value from the older household member	26
9	None of the above	Use the maximum of the values from side A, side B, the screener, and 2	28

¹ In households where a pair of respondents was selected, the household size had to be greater than or equal to two.

14.5.1.2 Household Composition Age Count Variables

It was logical that the ages of other household members would be good covariates for imputation of pair relationships. Such variables were also important for imputing multiplicity and household-level domain counts. The household-consistent age counts were limited to the following age ranges: younger than 12, 12 to 14, 15 to 17, 18 to 20, 21 to 25, 26 to 34, 35 to 49, and 50 or older. These variables were called AGE011, AGE1214, AGE1517, AGE1820, AGE2125, AGE2634, AGE3549, and AGE50P, respectively.

The first step in this process was to count the nonmissing ages for roster members in the household for each pair member. Sometimes it was necessary to adjust the count because the ages could not be matched exactly. For example, suppose a 38-year-old and a 17-year-old were selected, and the 17-year-old responded first. Suppose also that the 17-year-old turned 18 before the 38-year-old responded. Hence, the 17-year-old would have had an age of 18 in the 38-year-old's roster. However, because the younger pair member was 17 at the time of survey completion, the ages of interest for this pair domain were defined to be 17 and 38. Hence, it was necessary to account for this by creating a new roster age variable that matched the age provided in the other pair member's questionnaire. The age counts using this new roster age variable were equivalent to subtracting 1 from the previously obtained 18-25 count and adding 1 to the previously obtained 12-17 count in the 38-year-old's roster. These adjustments were made whenever a match was made between one pair member's roster and another pair member's questionnaire age and gender and the ages did not match exactly.

If no roster ages were missing, the sum of these counts was equal to the imputed household size IRHHSIZE. The reported household size was not considered here, because the counts were obtained from an edited roster. As with household size, a series of priority conditions was used to obtain the most likely count within each age group. If the appropriate count was ambiguous because of disagreement between the pair members, the quality of the roster and the age of the respondent (in that order) were used to determine the appropriate count. The roster quality was determined by the number of bad or missing roster entries and the quality of the match between the respondent's roster and the other pair member's questionnaire age and gender. If only one household member was selected as a respondent, counts from the screener roster were used to fill in bad or missing age counts. The conditions used to reconcile the age count variables are detailed in [Table 14.13](#). If the age counts could not be determined using the conditions, missing counts were filled in using subtraction, where possible. Any remaining missing counts were left blank. If any of the age counts were still missing after reconciliation, then only HHSIZE was used in the imputation models.

Table 14.13 Conditions Used to Reconcile the Age Counts across Pairs

Condition Number	Description	Decision	Number of Occurrences ¹
1	All counts match for sides A and B and the sum is equal to HHSIZE	Use values from side A	11,364 ²
2	The sum of the counts for side A is equal to HHSIZE, but the sum of the counts for side B is not	Use values from side A	1,272
3	The sum of the counts for both sides is equal to HHSIZE, but only the counts from side A match the modified age counts ³	Use values from side A	56
4	Both sides are consistent with HHSIZE and the modified age counts, but only the counts from side A match the screener	Use values from side A	483

Table 14.13 Conditions Used to Reconcile the Age Counts across Pairs (continued)

Condition Number	Description	Decision	Number of Occurrences ¹
5	Both sides are consistent with HHSIZE, but neither matches the screener; the roster on side A had a better quality match to the other pair member's age and gender	Use values from side A	44
6	Neither side had a good quality match to the other pair member; respondent A is older	Use values from side A	111
7	HHSIZE is equal to TGOODAGA	Use values from side A	2
8	HHSIZE is equal to the screener household size	Use values from the screener	24
9	None of the above	Leave blank	0 ⁴

¹ Most of the time, the same priority condition was used for every age group. Differences are noted below.

² For AGE011, this count was 11,362; for AGE2634 and AGE50p, this count was 11,363; for AGE3549, this count was 11,361; and for AGE15p, this count was 11,360.

³ See Section 14.5.1.1.

⁴ For AGE011, this count was 1; for AGE3549, this count was 2; and for AGE15p, this count was 3.

14.5.1.3 Household Composition Age Counts of Males

For some pair variables, particularly spouse-spouse pairs, knowledge of the gender of the roster member was important for imputation purposes. In a similar manner to that used in the creation of the household composition age count variables, variables counting the number of males within the given age ranges were created. Disagreements between pair members were resolved in a similar manner to what was done with the household composition age count variables. The names of the male age counts were MALE011, MALE1214, MALE1517, MALE1217, MALE1220, MALE1825, MALE2634, MALE3549, and MALE50P.

14.5.2 Stage One Imputation: Pair Relationships

Missing pair relationships were filled in using the single RP/single PRD type of PMN. In this stage, the imputation set involved only the edited variable PAIRREL, with its corresponding imputation-revised variable IRPRREL. Because the pair relationship varies according to the ages of the respondents, modeling and imputation were done independently within each of 11 age group pairs. [Table 14.14](#) presents these age group pairs, as well as the pair relationships prevalent within each age group pair.

Table 14.14 Age Group Pairs with Associated Possible Pair Relationships

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence) ¹	
		≥ 10% Prevalence ²	< 10% Prevalence
0	12-14/12-14	Sibling-sibling	Other relationship
1	12-14/15-17	Sibling-sibling	Other relationship
2	12-14/18-25	Sibling-sibling	Other relationship; parent-child; spouse-spouse**

Table 14.14 Age Group Pairs with Associated Possible Pair Relationships (continued)

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence) ¹	
		≥ 10% Prevalence ²	< 10% Prevalence
3	15-17/15-17	Sibling-sibling	Other relationship; spouse-spouse**
4	15-17/18-25	Sibling-sibling	Other relationship; spouse-spouse*; parent-child*
5	18-20/18-25	Sibling-sibling; other relationship; spouse-spouse	Parent-child**
6	21-25/21-25	Spouse-spouse; other relationship; sibling-sibling	Parent-child**
7	12-14/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling
8	15-17/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling; spouse-spouse*
9	18-20/26+	Parent-child; other relationship	Sibling-sibling; grandparent-grandchild; spouse-spouse*
10	21+/26+	Spouse-spouse; parent-child; other relationship	Sibling-sibling; grandparent-grandchild

*Pair relationship occurs in less than 1 percent of the overall total number of pair relationships.

**The pair relationship is so rare that it does not appear in the age group pair in every survey year.

¹ The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships. The spouse-spouse domain as listed here actually consists of two domains (spouse-spouse-with-children and spouse-spouse-without-children) that have been collapsed for the purposes of making the table easier to read. "Other relationship" refers to a relationship other than sibling-sibling, parent-child, grandparent-grandchild, or spouse-spouse.

² The pair relationships each form at least 10 percent of the overall total number of pair relationships within the given age group pair, and the total is at least 85 percent of the overall total.

14.5.2.1 Response Propensity Step

For a respondent pair to be considered complete, the pair relationship must be definitively established. In terms of the variable PAIRREL, this meant that the pair had to have a value of PAIRREL within the range of 1 to 9 or = 11 or 12. A value of PAIRREL = 13 was considered complete, even though the pair relationship was not definitively established, because it was known that the pair relationship was not a relationship of interest. RP adjustments then were calculated for each age group pair to make the respondent pair weights representative of the entire sample of pairs. Because the modeling of the final pair weight adjustments was not completed at the time of the pair imputations, the pair-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment. These adjustments were calculated using an item RP model, which is a special circumstance of the generalized exponential model. See Appendix A of the *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 12: Questionnaire Dwelling Unit-Level and Person Pair-Level Sampling Weight Calibration* report (Center for Behavioral Health Statistics and Quality, 2025c) for technical details about this procedure.

14.5.2.2 Prediction Step

After the weights were adjusted using the item RP model within each age group pair, logistic regression models were fit using the adjusted weights. Preferred covariates in these models included the age count variables described in Section 14.5.1.2. However, variables with missing values for some observations cannot be used as covariates in the models. To allow for better models whenever possible, the data were partitioned into two groups: those with nonmissing values for all age count variables, and those where one or more values of the age count variables were missing. When all of the age count variables were nonmissing, the predicted mean model was fit using these variables. Otherwise, the model was fit using the overall household count. This resulted in two predicted mean models for some of the 11 age group pairs.¹⁹⁷ These models were combined again before the hot-deck imputation step.

All modeling was done using the SUDAAN® (RTI International, 2013) procedure MULTILOG; however, the number of levels in the response variable varied based on the age group pair.¹⁹⁸ For age group pairs 0 through 4 and 7 through 9, dichotomous logistic regression models were built. Because there were three outcomes with age group pairs 5, 6, and 10, polytomous logistic models were fit for these age group pairs. All the models incorporated weights and were calibrated to account for item nonresponse (where a pair responded to the survey but the pair relationship was unknown) using the item RP models described in Section 14.5.2.1.

Ideally, each type of pair relationship within an age group pair would constitute a response category in a multinomial response model. However, the number of occurrences of some pair relationships within each age group pair was very small. Hence, it was not feasible to fit multinomial models that cover all the possible pair relationships for a given age group pair. Therefore, in the modeling step, some of the response categories were combined with separate assignments of imputed values within each of the 11 age group pairs. Priority was given to placing the pair relationships “of interest” into separate categories. Sometimes pair relationships that were not of interest were combined with other categories, even if there were sufficient numbers to have a separate category in the multinomial model. [Table 14.15](#) presents the response categories that were used for modeling. The delineation between categories that were combined for modeling was determined in the hot-deck step.

As an example, consider age group pair 5. In this age group pair, there are typically four types of pair relationships that have a sufficient number of respondent pairs to fit an adequate model, including both spouse-spouse domains, sibling-sibling pairs, and all others. Models with fewer response levels are generally easier to fit because there are more observations in each response level. Because only two of those four were pair relationships of interest, the two spouse-spouse domains were used as levels in the response variable. The third level was obtained by

¹⁹⁷ Partitioning the observations into two groups is similar to including an indicator variable in the PRD model. Though two models were fit in practice, in theory this is similar to fitting only one PRD model. Thus, this is still considered the single RP/single PRD type of PMN.

¹⁹⁸ Though the SUDAAN procedure RLOGIST could have been used for all age group pairs except 5, 6, and 9, MULTILOG was used for all models for coding simplicity.

combining the sibling-sibling and other relationship pairs. There are typically a small number of parent-child pairs, which also were combined with the other relationship pairs.

Table 14.15 Modeled Pair Relationships within Age Group Pairs

Age Group Pair Number	Age Group Pair	Number of Levels in Response	Levels of Modeled Response
0	12-14/12-14	2	Sibling-sibling; all others
1	12-14/15-17	2	Sibling-sibling; all others
2	12-14/18-25	2	Sibling-sibling; all others
3	15-17/15-17	2	Sibling-sibling; all others
4	15-17/18-25	2	Sibling-sibling; all others
5	18-20/18-25	3	Both spouse-spouse pair relationships; ¹ all others
6	21-25/21-25	3	Both spouse-spouse pair relationships; ¹ all others
7	12-14/26+	2	Parent-child; all others
8	15-17/26+	2	Parent-child; all others
9	18-20/26+	2	Parent-child; all others
10	21+/26+	3	Both spouse-spouse pair relationships; ¹ all others

¹ The two spouse-spouse pair relationships are spouse-spouse and spouse-spouse with children younger than 18. The pair relationships labeled "spouse-spouse" include partner-partner pair relationships.

14.5.2.3 Hot-Deck Step

Likeness constraints used in imputation of the pair relationship were generally based on the number of household members in various age groups and on the marital status and genders of the respondents. Logical constraints were limited to the information that was already known about the pair, as denoted by the level of the variable PAIRREL. For example, if PAIRREL = 14, then no information was available about the identity of the pair relationship and no logical constraint was needed. On the other hand, if PAIRREL = 15, this meant that the pair relationship was either a parent-child pair where the child was aged 12 to 14 or it was some relationship other than spouse-spouse, parent-child, grandparent-grandchild, or sibling-sibling.

14.5.3 Stage Two Imputation: Multiplicity Counts

Often, when the pair relationships were not defined, multiplicity counts also were not defined. In addition, there were a handful of times where multiplicity counts could not be determined, even when the pair relationship was known. In all of these circumstances, imputation was required to determine the multiplicity count, and the single RP/single PRD type of PMN was used for imputation.

The multiplicity count variables were divided into six imputation sets: four sets for the multiplicities associated with the four sibling-sibling pair domains; one set for the parent-child, child focus domains; and one set for the parent-child, parent focus domains. The variables in each imputation set are provided in [Table 14.16](#). Because the parent-child counts are

hierarchical (i.e., the count for 12-17 must be less than or equal to the count for 12-20), only the counts for the 12-20 age group were modeled. Using the predicted means from these models, a single donor pair for each focus was selected from which the multiplicity counts were determined for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs. No imputation was required for the spouse-spouse multiplicity counts because a selected respondent in a spouse-spouse pair was assumed to have had only one spouse.

Table 14.16 Modeled Multiplicities within Domains

Imputation Set	Domain	Model Type	Base Variables	Imputation-Revised Variables
1	Parent-Child, Child Focus	Logistic	MCPCC14, MCPCC57, MCPCC17, MCPCC20	IRMPCC14, IRMPCC57, IRMPCC17, IRMPCC20
2	Parent-Child, Parent Focus	Poisson	MCPCP14, MCPCP57, MCPCP17, MCPCP20	IRMPCP14, IRMPCP57, IRMPCP17, IRMPCP20
3	Sibling-Sibling (12-14/15-17), Older Sibling Focus	Poisson	MCS1417	IRMS1417
4	Sibling-Sibling (12-14/15-17), Younger Sibling Focus	Poisson	MCS1714	IRMS1714
5	Sibling-Sibling (12-17/18-25), Older Sibling Focus	Poisson	MCS1725	IRMS1725
6	Sibling-Sibling (12-17/18-25), Younger Sibling Focus	Poisson	MCS2517	IRMS2517

14.5.3.1 Imputation for Parent-Child Multiplicity, Child Focus (Imputation Set 1)

The first imputation set included the four parent-child, child focus multiplicity variables. Modeling was done only for the parent-child (12-20) variable, and multivariate assignment of all four variables was done in the hot-deck step.

14.5.3.1.1 Response Propensity Step

For a respondent pair to be considered complete regarding the parent-child multiplicities, the multiplicity had to be nonmissing for the domains with children aged 12 to 20. A nonmissing multiplicity for this domain would automatically guarantee nonmissing multiplicities for the subset parent-child domains. RP adjustments were then calculated to make the respondent pair weights representative of the entire sample of pairs.

14.5.3.1.2 Prediction Step

For the child-focus parent-child domains, the count being modeled was the number of parents of children aged 12 to 20 who were part of a parent-child pair. Most of the time, only two responses were possible: one parent or two parents. There were rare instances where the child could have three parents living in the household, perhaps due to some combination of biological, step, foster, or adoptive parents. For the purposes of modeling, these instances were collapsed with the two-parent households. Similar to the procedure described in Section

14.5.2.2, the data were divided into two groups based on whether the age count variables were nonmissing, and separate models were fit for each group. The fit models were binomial logistic regression models using the SUDAAN procedure RLOGIST.

14.5.3.1.3 Hot-Deck Step

Though modeling was only done for the 12-20 age group, the multiplicity counts for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs were assigned from a single donor pair. Likeness constraints were generally based on the pair relationship, the household size, and the number of household members in various age groups. These constraints are described in detail in [Table E.246](#) in Appendix E. There were no logical constraints.

14.5.3.2 Imputation for All Other Multiplicities (Imputation Sets 2 through 6)

The imputation process for the other responses (parent-focus parent-child and sibling-sibling multiplicity counts) was similar to that for the child-focus parent-child counts. The main difference is in the PRD step where Poisson regression models were used to model the counts. The counts of the number of children or siblings were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance.

As with the child-focus parent-child multiplicity counts, the parent-focus parent-child multiplicity counts were only modeled for the 12-20 age group and counts from all four age groups were assigned from a single donor pair. For the sibling-sibling multiplicity counts, there was only one variable in each imputation set, and each count was modeled separately.

14.5.4 Stage Three Imputation: Household-Level Person Counts

Because of the difficulty in definitively determining household-level counts, imputation was not uncommon. Household-level person counts were divided into five imputation sets based on the domains listed in Section 14.4: one for each of the two sibling-sibling counts, one for the spouse-spouse counts, one for the spouse-spouse with children counts, and one for the parent-child counts. The first four imputation sets were handled using the single RP/single PRD type of PMN, whereas the imputation of the parent-child household counts was done using the multiple RP/multiple PRD type of PMN. For these counts, separate models were fit for the child-focus and parent-focus counts, and the predicted means from both models were brought together in one hot-deck step.

As with the multiplicities, the parent-child domains were hierarchical, so the imputations could not have been conducted independently if consistency was to be maintained. Hence, models were fit only to the parent-child domains for the 12-20 age group, and the household-level person counts were assigned for the 12-14, 12-17, and 12-20 parent-child pair domains from a single donor. The household-level person counts for the 15-17 parent-child domains were not determined as they can be easily derived. The spouse-spouse household-level person counts were also hierarchical in that knowledge of whether a spouse-spouse pair was in the household was required before one could say that the pair had children. Therefore, imputations of the

spouse-spouse counts were processed first, followed by imputations of whether the spouse-spouse pairs in the household had children.

Household-level person counts were defined for all respondents, regardless of which pair they belonged to, or even whether they were within a pair at all.

14.5.4.1 Imputation for Sibling-Sibling (12-14/15-17), Older Focus Household Counts (Imputation Set 1)

The first imputation set included one base variable, HCS1417, and the corresponding imputation-revised variable IRHS1417. The imputation process for this variable is detailed in Sections 14.5.4.1.1 through 14.5.4.1.3.

14.5.4.1.1 Response Propensity Step

For a pair or single respondent to be considered complete, the household-level person counts had to be nonmissing for all the variables being imputed. In addition to being separated into pair and nonpair households, respondents also were split by age for both the modeling of the RP and the predicted means. For the pairs, households where both pair members were younger than 18 were placed in one group, and the remaining pairs were in the other age group. For the single respondents, one age group consisted of respondents who were younger than 18, and the other consisted of respondents who were 18 or older. For pairs, RP adjustments were calculated to make the household weights representative of the entire sample of pairs. For single respondents, household weights also were used, and the adjustments were calculated to make the respondent household weights representative of the entire sample of households that were not part of a pair.

14.5.4.1.2 Prediction Step

After the RP adjustment was complete, the data were split into two groups based on whether the household-level age count variables were nonmissing, as described in Section 14.5.2.2. The outcome variable for the household-level sibling-sibling (12-14/15-17) count models was the number of household members aged 15 to 17 with a sibling aged 12 to 14 living with them. These counts could have a value of zero, which distinguished them from the multiplicities from a modeling point of view. Poisson regression was used to fit the models for the household-level person counts corresponding to the sibling-sibling domains. The data were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance. Modeling was done using the SUDAAN procedure LOGLINK.

14.5.4.1.3 Hot-Deck Step

After the modeling steps were complete, the two age groups were combined for one hot-deck imputation step. Imputation was conducted together for pair and nonpair households. Likeness constraints used in the imputation of household counts were generally based on the number of families in the household, the household size, and the number of household members in various age groups. Logical constraints were based on the bounds created during the editing process. These constraints are described in detail in [Tables E.266](#) and [E.267](#).

In those instances where an imputed value could not be found after loosening all the likeness constraints, the imputed value was determined by doing a random imputation within bounds derived from the household composition.

14.5.4.2 Imputation for Sibling-Sibling (12-17/18-25) and Both Spouse-Spouse Household Counts (Imputation Sets 2 through 4)

The second, third, and fourth imputation sets were processed similarly to the first. One large difference was that when modeling the spouse-spouse-with-children counts, the data were not separated according to age groups. This applied to both the RP adjustments and the calculation of predicted means. Because of the hierarchical relationship between the spouse-spouse and the spouse-spouse-with-children counts, the RP adjustment for the spouse-spouse-with-children domain adjusted the weights to be representative of all spouse-spouse pairs rather than the entire sample. Missing counts for the spouse-spouse-with-children domain were not replaced via imputation until it was known definitively, after the hot-deck step of the PMN imputation, whether a household had spouse-spouse pairs.

Multinomial logistic regression was used to model the count of spouse-spouse pairs, with the possible responses being zero, one, and two or more spouse-spouse pairs in the household. Whether or not the spouse-spouse pairs had children younger than 18 was modeled with binomial logistic regression. These models were fit using the MULTILOG procedure in SUDAAN.

Sometimes two family units were in a household. In these situations, the predicted means were set to zero and were grouped together using a likeness constraint during hot-deck imputation.

14.5.4.3 Imputation for Parent-Child Household Counts (Imputation Set 5)

In contrast to the first four imputation sets, the parent-child household counts were imputed using the multiple RP/multiple PRD type of PMN. This imputation set included the three child-focus counts, HCPCC14, HCPCC17, and HCPCC20, as well as the three parent-focus counts, HCPCP14, HCPCP17, and HCPCP20. The corresponding imputation-revised variables are IRHPCC14, IRHPCC17, IRHPCC20, IRHPCP14, IRHPCP17, and IRHPCP20, respectively. The child-focus and parent-focus counts were modeled separately and combined in one final hot-deck step. Unlike most imputation sets processed using the multiple RP/multiple PRD type of PMN, however, no provisional hot-deck steps were implemented. Because the child-focus counts were not used as covariates for the parent-focus counts and the parent-focus counts were not used as covariates for the child-focus counts, no provisional hot-deck steps were necessary, and the RP and PRD steps for the child-focus counts were run in parallel with the RP and PRD steps for the parent-focus counts.

14.5.4.3.1 First Response Propensity Step: Child Focus

All respondents were in the domain for all household counts. For a pair or single respondent to be considered complete, the three child-focus counts had to be nonmissing. As with the processing of the sibling-sibling counts and the spouse-spouse counts, the data were divided into two age groups based on whether the respondent or pair of respondents was 18 or older.

14.5.4.3.2 First Prediction Step: Child Focus

For the child-focus counts, the modeled response was the number of children aged 12 to 20 in the household with at least one parent living with them. This was modeled using Poisson regression, where the data were scaled using the observed variance to account for underdispersion. The LOGLINK procedure in SUDAAN was used to fit the model. The predicted mean used in the final hot-deck step was the predicted number of children aged 12 to 20 in the household with at least one parent living with them.

14.5.4.3.3 Second Response Propensity Step: Parent Focus

This RP step was identical to the RP step for the child-focus counts, except that item respondents were those whose three parent-focus counts were nonmissing.

14.5.4.3.4 Second Prediction Step: Parent Focus

For the parent-focus counts, the modeled response was a three-level variable, based on the number of parents in the household with children aged 12 to 20: zero, one, or two or more. Polytomous logistic regression was used to fit the model as implemented by the MULTLOG procedure in SUDAAN. The predicted means used in the final hot-deck step were the predicted probabilities associated with each of the three levels of the response variable.

14.5.4.3.5 Hot-Deck Step

The predicted means from the models fit for the child-focus and parent-focus counts were brought together in one hot-deck imputation step. Because there are six variables in the imputation set, and any combination of them could be missing, there were 63 possible missingness patterns. These patterns are enumerated in [Table E.265](#). As with the other four imputation sets, logical constraints were based on the bounds created during the editing process, and likeness constraints were generally based on the number of families in the household, the household size, and the number of household members in various age groups. These constraints are described in detail in [Tables E.263](#) and [E.264](#).

When there were two family units in the household, donors for two-family households were required to be from other two-family households. Because there were very few complete two-family households, imputation was frequently random between the bounds.

References

- Allison, P. D. (2008a). *Convergence failures in logistic regression*. Paper presented at the SAS Global Forum 2008, Statistics and Data Analysis.
<https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/360-2008.pdf>
- Allison, P. D. (2008b). Missing data. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE handbook of quantitative methods in psychology*. Sage Publications, Inc.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.).
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>
- Center for Behavioral Health Statistics and Quality. (2014). *National Survey on Drug Use and Health: 2012 Questionnaire field test final report*. <https://www.samhsa.gov/data/report/nsduh-2012-questionnaire-field-test-report>
- Center for Behavioral Health Statistics and Quality. (2017). *Evaluation of imputation methods for the National Survey on Drug Use and Health*.
<https://www.samhsa.gov/data/report/evaluation-imputation-methods-national-survey-drug-use-and-health>
- Center for Behavioral Health Statistics and Quality. (2018). *2016 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 10: Editing and imputation report*.
<https://www.samhsa.gov/data/report/2016-editing-and-imputation-report>
- Center for Behavioral Health Statistics and Quality. (2021). *2019 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 10: Editing and imputation report*.
<https://www.samhsa.gov/data/report/nsduh-2019-editing-and-imputation-report>
- Center for Behavioral Health Statistics and Quality. (2023a). *Key substance use and mental health indicators in the United States: Results from the 2022 National Survey on Drug Use and Health*. <https://www.samhsa.gov/data/report/2022-nsduh-annual-national-report>
- Center for Behavioral Health Statistics and Quality. (2023b). *Results from the 2022 National Survey on Drug Use and Health: Detailed tables*. <https://www.samhsa.gov/data/report/2022-nsduh-detailed-tables>
- Center for Behavioral Health Statistics and Quality. (2023c). *2022 National Survey on Drug Use and Health: Editing and imputation evaluation report* (unpublished internal documentation). Substance Abuse and Mental Health Services Administration.
- Center for Behavioral Health Statistics and Quality. (2024a). *2023 National Survey on Drug Use and Health (NSDUH): Methodological summary and definitions*.
<https://www.samhsa.gov/data/report/2023-methodological-summary-and-definitions>

Center for Behavioral Health Statistics and Quality. (2024b). *2023 National Survey on Drug Use and Health public use file codebook*. <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/datafiles>

Center for Behavioral Health Statistics and Quality. (2024c). *Results from the 2023 National Survey on Drug Use and Health: Detailed tables*. <https://www.samhsa.gov/data/report/2023-nsduh-detailed-tables>

Center for Behavioral Health Statistics and Quality. (2025a). *2023 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 2: Sample design report*. <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/methodology>

Center for Behavioral Health Statistics and Quality. (2025b). *2023 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 11: Person-level sampling weight calibration*. <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/methodology>

Center for Behavioral Health Statistics and Quality. (2025c). *2023 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 12: Questionnaire dwelling unit-level and person pair-level sampling weight calibration*. <https://www.samhsa.gov/data/data-we-collect/nsduh-national-survey-drug-use-and-health/methodology>

Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA*. <https://www.amstat.org/membership/asa-communities/sections-amp-interest-groups>

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the 1980 American Statistical Association, Survey Research Methods Section, Houston, TX* (pp. 721-726). <https://www.amstat.org/membership/asa-communities/sections-amp-interest-groups>

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

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

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, 33(2), 357-362.

Iannacchione, V. (1982). Weighted sequential hot deck imputation macros. In *Proceedings of the Seventh Annual SAS Users Group International Conference* (pp. 759-763). SAS.

- 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(2), 184-189.
- Kott, P. S., & Folsom, R. E. (2010). Weights, double protection, and multiple imputation. In *Proceedings of the Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods* (pp. 431-440). <https://www.amstat.org/membership/asa-communities/sections-amp-interest-groups> 
- Kroutil, L., & Myers, L. (2002). Development of editing rules for CAI substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 85-109). Substance Abuse and Mental Health Services Administration.
- Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. John Wiley.
- Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research*, 19(Suppl. 1), 49-60.
- Office of Applied Studies. (2009). *Results from the 2008 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Substance Abuse and Mental Health Services Administration.
- Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register*, 62(210), 58781-58790.
- Office of Management and Budget. (2009). *Update of statistical area definitions and guidance on their uses* (OMB Bulletin No. 10-02). The White House. <https://www.whitehouse.gov/omb/information-for-agencies/bulletins/#2009-2016>
- Raghunathan, T. E., Solenberger, P. W., & Van Hoewyk, J. (2002). *IVEware: Imputation and variance estimation software user guide*. University of Michigan, Survey Methodology Program, Survey Research Center, Institute for Social Research.
- Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research*, 8(2), 110-123.
- RTI International. (2013). *SUDAAN® language manual, Release 11.0.1*.
- Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics*, 4(1), 87-94.
- SAS Institute Inc. (2014). *SAS/STAT® 13.2 user's guide: High-performance procedures*.
- Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). Chapman and Hall/CRC.

Shiffman, S., Hickcox, M., Gnys, M., Paty, J. A., & Kassel, J. D. (1995, March). *The Nicotine Dependence Syndrome Scale: Development of a new measure*. Poster presented at the annual meeting of the Society for Research on Nicotine and Tobacco, San Diego, CA.

Shiffman, S., Paty, J. A., Kassel, J. D., Gnys, M., & Zettler-Segal, M. (1994). Smoking behavior and smoking history of tobacco chippers. *Experimental and Clinical Psychopharmacology*, 2(2), 126.

Shiffman, S., Waters, A. J., & Hickcox, M. (2003). *The Nicotine Dependence Syndrome Scale: A multi-dimensional measure of nicotine dependence*. Unpublished manuscript.

Singh, A., Grau, E., & Folsom, R. (2004). Imputation and unbiased estimation: Use of centered predictive mean neighborhoods method. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* (pp. 4351-4358). <https://www.amstat.org/membership/asa-communities/sections-and-interest-groups> ↗

Acknowledgments

This National Survey on Drug Use and Health (NSDUH) report was prepared by the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality, and by RTI International. Work by RTI was performed under contract number 75S20322C00001. Marlon Daniel and Carlos Graham served as contracting officer representatives, and David Hunter served as RTI project director.

Contributors at RTI included Tricia Adams, Kristen Gullledge Brown, Lisa Carpenter, Chuchun Chien, Susan Edwards, Barbara Felts, Emily Gentles, Erica Hirsch, Larry Kroutil, Peilan Chen Martin, Andrew Moore, Michael Penne, Victoria Scott, and Chelsy Thompson.

Special thanks are due to several current and former RTI staff members. Ralph Folsom, Eric Grau, and Avinash Singh (all formerly with RTI) co-developed the predictive mean neighborhood (PMN) methodology, and Ralph Folsom and Phillip Kott co-developed the modified predictive mean neighborhood (modPMN) methodology.

Appendix A: Sample Design Overview and Related Household-Level and Person-Level Files

A.1 Overview of the Sample Design

For the 2023 administration of the National Survey on Drug Use and Health (NSDUH), a person was randomly selected for an interview through a five-stage sample selection process. States were first stratified into a total of 750 state sampling (SS) regions. Within each of these SS regions, a sample of census tracts was selected (i.e., the first stage of selection) with probabilities proportional to a composite size measure and with minimum replacement. For the second stage of selection, adjacent census block groups within selected census tracts were aggregated to meet the minimum dwelling unit (DU) requirements. One census block group was then selected per sampled census tract with probability proportionate to a composite size measure and with minimum replacement. The selection of census block groups at the second stage of selection allowed a transition to a hybrid field enumeration and address-based sampling (ABS) frame for the 2022 and 2023 NSDUHs.

For the 2023 NSDUH, the hybrid field enumeration and ABS approach differed for the portion of the sample that was carried over from the 2022 NSDUH and for the portion that was new in 2023.¹ For both portions, census block groups selected at the second stage of selection were first evaluated using a set of ABS coverage criteria. Those census block groups that met all ABS coverage criteria served as ABS segments. For the new portion of the 2023 sample, the census block group was field enumerated if it did not meet all ABS coverage criteria. For the portion of the sample that was carried over from the 2022 NSDUH, a smaller geographic area was selected for field enumeration for cost-efficiency. In these census block groups, adjacent census blocks were combined to form third-stage sampling units or smaller area segments, and one smaller area segment was selected with probability proportional to population size. For the 2023 NSDUH, the “segment” is the geographic area for which the DU frame is constructed, even though some segments are second-stage sampling units and other segments are third-stage sampling units.

Once the sample segments, including both ABS and field enumeration segments, were selected, lists of DUs were constructed. For ABS segments, the DU list consisted of addresses on the ABS frame that geocoded into the sampled area. For field enumeration segments, specially trained staff visited areas and created lists of all eligible DUs within the sample segment boundaries. The ABS and field enumerated lists served as the frames for the fourth stage of sample selection. After the DUs were selected within each segment, a roster of all people aged 12 or older was obtained from an eligible member of the selected DU.

The 2023 NSDUH used a combination of web and in-person data collection methods (i.e., multimode data collection). At the beginning of each calendar quarter, a large sample of DUs were mailed an invitation to participate via the web. Several days later, field interviewers

¹ The NSDUH sample design includes a 50 percent overlap in sampled areas between consecutive survey years.

began following up with pending DUs for in-person data collection. The *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 2: Sample Design Report* (Center for Behavioral Health Statistics and Quality, 2025a) contains more information regarding the sample design.

At the end of the survey year, a household-level file and a person-level file were created to record the information obtained from the sampling processes described above. The household-level and person-level files were used in the final creation of the person-level and pair-level analysis weights. In addition, the person-level file was later subset into a smaller data file that contained only respondents who were considered “completed” interviews; this file was used for analysis. Refer to Section A.3 for the definition of a completed interview.

A.2 Dwelling Unit-Level Eligibility and Completeness Criteria

Before proceeding with the fifth stage of sample selection (i.e., within selected households), a set of rules was used to determine whether a DU was eligible to be selected. Eligibility of the DU was recorded in the binary variable DUELIG, where a value of 1 indicated eligibility. When data were collected in person, interviewers could identify ineligible DUs to allow the ineligible DUs to be removed from the sample before weighting. Two examples of DUs deemed to be ineligible included those defined as “vacant” and those determined to be “not a primary residence.” For web-based data collection, DU members needed to initiate the screening process for eligibility to be confirmed. If an adult DU member did not initiate the web-based screening interview and the DU was not visited by a field interviewer, the DU’s eligibility status was unknown. For these DUs, the eligibility status was randomly imputed according to the historic DU eligibility rate in the state where the DU was located. DUs that were imputed to be ineligible were excluded from the weighting process before the screening DU nonresponse adjustment.

Occasionally, DU members were eligible but failed to complete the screening process. Reasons for not completing the screening process were recorded, including situations such as “language barrier,” “refusal,” and “denied access.” Completeness of the screening process for the DU was recorded in the binary variable DUCOMP, where a value of 1 indicated completeness.

During the third stage of sampling, it was possible to select a sample segment more than once because samples were selected with replacement. These duplicated segments had different segment IDs (SEGIDs) for each duplicate. To avoid selecting the same DU twice at the fourth stage of selection, the sample was selected from one SEGID. Then the sample DUs were split as evenly as possible between the two SEGIDs, and this information was updated in the household-level and person-level files.

A.3 Person-Level Eligibility and Completeness Criteria

During screening, respondents were asked to identify all eligible household members so that only eligible individuals were listed and, therefore, potentially selected. Eligible respondents were permanent residents of the DU who were noninstitutionalized civilians aged 12 years or older at the time of the screening. Ineligibles included the following: not a permanent resident in the DU, younger than 12 years old, person in the military, and roster error. Eligibility was

recorded in the binary variable PRELIG, which had a value of 1 if the household member was eligible. Respondents who were selected were recorded in the binary variable PRSEL, where 1 indicated a selected individual. It was possible for an individual to be selected, but at the time of the interview, to be determined ineligible. Examples of changes from eligibility to ineligibility included “the selected person turned out not to be a permanent resident in the DU” and “roster error.” If this occurred, the value of PRELIG was changed from 1 to 0.

A summary of the number of selected, eligible, and completed DUs is shown in [Table A.1](#). The number of eligible, selected, and interviewed individuals also is summarized in the table. Due to the expectation that web response rates would be lower than in-person response rates, a very large sample was selected and released for web data collection.

Table A.1 NSDUH Selected, Eligible, and Interviewed Households and People: 2023

Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings (Dwelling Units)	Eligible People	Selected People	Interviewed Respondents	Completed Interviews
874,764	804,971	198,246	427,464	135,737	68,384	67,679

To be considered a completed interview for purposes of analysis, a respondent had to provide “yes” or “no” answers to the cigarette usage gate question and to at least 9 of the following additional drug usage gate questions: (1) chewing tobacco, (2) snuff, (3) cigars, (4) alcohol, (5) marijuana, (6) cocaine (in any form), (7) heroin, (8) hallucinogens, (9) inhalants, (10) pain relievers, (11) tranquilizers, (12) stimulants, and (13) sedatives.² Completeness of the survey for eligible individuals was recorded in the binary variable PRCOMP, which had a value of 1 if the respondent was a completed interview, and 0 if not. For a summary of the number of completed interviews in the 2023 survey, see [Table A.1](#).

A.4 Variables in the Household-Level and Person-Level Files

This section documents some of the important person-level variables that were created for the household-level and person-level files.

Screeners-level demographic variables were created from the screener roster information in the household-level and person-level files. XAGE was the screener age, which either could be “continuous” (single-year ages) or categorical. An in-person respondent could choose to give an age category instead of the actual age. The age categories with their accompanying codes were 199 = 12 to 17 years old; 299 = 18 to 25 years old; 399 = 26 to 34 years old; 499 = 35 to 49 years old; and 599 = 50 years old or older. Screener race (XRACE1-XRACE5), screener Hispanicity (XHISP), screener sex at birth (XSEX), and screener gender identity (XGID1-XGID4) also were produced from the screener roster information. XRACE1 through XRACE5 were indicator variables representing White, Black or African American, American Indian or Alaska

² For more details on the “usable interview rule,” see Section 2.2.3.

Native, Asian, and Native Hawaiian or Other Pacific Islander, respectively. XSEX indicated either Male or Female. XGID1 through XGID4 were indicator variables representing Male, Female, Transgender, or Different Term. The household-level variable PAIRSEL represented the number of people within each age group selected from a DU. It was a 20-level variable indicating whether zero, one, or two individuals were selected in the household from the five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). If two people were selected from the household, this variable indicated the age groups of both pair members. Similar to PAIRSEL, the household-level variable PAIRRESP had 20 levels, which indicated whether zero, one, or two people completed the interviews from the five age groups within a household.

As described in the 2023 Sample Design Report (Center for Behavioral Health Statistics and Quality, 2025a), states were partitioned into SS regions, which were further partitioned into clusters of adjacent block groups or blocks called “segments.” The variable SEGID (segment ID number) was a two-letter state abbreviation followed by a two-digit SS region and a two-letter segment identifier, with the fifth character indicating the survey year and the sixth character being the initial quarter assignment of the segment. Census region (REGION) was a four-level geographic variable recoded from the respondent’s state of residence. The four levels were Northeast, Midwest, South, and West. The population density variables PDEN10 and PDEN210³ classified respondents according to their living situation, whether it was in a rural or urban area, and, if urban, the size of the urban area. They were used to categorize segments where the respondents lived based on 2010 census data and the December 2009 Core-Based Statistical Area (CBSA) classifications.⁴ PDEN210 had five levels: segment in CBSA with 1 million or more people; segment in CBSA with 250,000 to 999,999 people; segment in CBSA with fewer than 250,000 people; segment in urban area but not in CBSA; and segment in rural area (not in CBSA and not in urban area). PDEN10 had three levels: segment in CBSA with 1 million or more people; segment in CBSA with fewer than 1 million people; and segment not in CBSA.

The variable STATE represented the Federal Information Processing Standards (FIPS) state codes for the 50 states and the District of Columbia. The variable STATE was created at the sampling stage and did not contain any missing values. The variables VESTR and VEREP were created to capture the sampling design structure. Each SS region appeared in a different variance estimation stratum (VESTR) every quarter. Two replicates (VEREP) were defined within each variance stratum. Each replicate consisted of four segments, one for each quarter of data collection. The segment-level variable RURURB10⁵ was derived from the 2010 census block-

³ PDEN10 and PDEN210 were similar to PDEN/PDEN2 and PDEN00/PDEN200 in previous sample designs. The suffix “10” indicates that the variables were derived from the 2010 census.

⁴ CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2009).

⁵ RURURB10 was similar to RURORURB and RURURB00 in previous sample designs. The suffix “10” indicates that this variable was derived from the 2010 census.

level designation of rural or urban.⁶ In the NSDUH sample, if 100 percent of the blocks were rural, the segment was defined as rural (RURURB10 = 1). If one or more of the blocks within a segment was urban, the segment was defined as urban (RURURB10 = 2). MODE captured the mode of response, either web (MODE = 1) or in person (MODE = 2). Other sampling variables such as DIVISION, SSREGION, GQTYPE, ID, STNAME, STUSAB, and QUARTER⁷ also were included in the household-level and person-level files.

⁶ The census classifies as urban all blocks located within urbanized areas (UA) and urban clusters (UC). UAs and UCs generally consist of core census block groups or blocks that have a population density of at least 1,000 people per square mile and surrounding census blocks that have an overall density of at least 500 people per square mile. In addition, under certain conditions, less densely settled territory may be part of each UA or UC.

⁷ For more details on these sampling variables, refer to the 2023 Sample Design Report (Center for Behavioral Health Statistics and Quality, 2025a).

Appendix B: Detailed Documentation of Edits for Variables from the Back-End Demographics, Substance Use, and Additional Sections in the 2023 NSDUH

Tables

B.1	Edits Pertaining to the Current Education Section.....	B-2
B.2	Edits Pertaining to the Employment and Workplace Section	B-7
B.3	Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use	B-9
B.4	Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine and Crack Cocaine	B-14
B.5	Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use.....	B-16
B.6	Edits Pertaining to Initiation Variables Based on Consistency Checks.....	B-18
B.7	Edits Pertaining to Initiation Variables Other Than for Parent/Child Relationships	B-20
B.8	Edits Pertaining to Parent/Child Initiation Variables	B-21
B.9	Miscellaneous Edits Pertaining to Tobacco and Nicotine Vaping Variables	B-25
B.10	Edits Pertaining to the Source of Prescription Psychotherapeutic Drug Variables	B-26
B.11	Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Section	B-28
B.12	Edits in the Special Drugs Section (Other Than Skip Patterns) Pertaining to Heroin Use and Needle Use.....	B-30
B.13	Edits in the Blunts Section Pertaining to Consistency Checks for Initiation	B-35
B.14	Edits in the Blunts Section Other Than Those Pertaining to Consistency Checks for Initiation.....	B-37
B.15	Edits in the Prior Substance Use Section Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables.....	B-41
B.16	Edits in the Prior Substance Use Section Pertaining to the Age, Year, and Month of Last Use (Other Than for Consistency Checks)	B-43
B.17	Edits Pertaining to the Alcohol and Drug Treatment Variables	B-47
B.18	Prioritization of Specific Skip Logic Edits in the Consumption of Alcohol Section	B-50
B.19	Edits in the Consumption of Alcohol Section Pertaining to the Binge Alcohol Use Variables	B-53
B.20	Edits Pertaining to the Special Topics Section	B-56
B.21	Edits Pertaining to the Mental Health Services Utilization Variables	B-57

B.1 Introduction

This appendix provides detailed documentation of edits that were implemented (where applicable) for variables from the back-end demographics, substance use, and additional sections that were self-administered for in-person interviews.¹ The remaining sections of the appendix are organized as follows:

- Section B.2 presents tables describing edits for selected back-end demographics variables corresponding to editing procedures that were discussed in Chapter 5.
- Section B.3 presents tables describing edits for substance use variables corresponding to editing procedures that were discussed in Chapter 6.
- Section B.4 presents tables describing edits for variables in additional sections corresponding to editing procedures that were discussed in Chapter 7.

The issues that are shown in the tables in these sections reflect changes to the questionnaire for 2023 (Section 1.4.1). Issues from prior years are no longer discussed if they have been resolved by questionnaire changes. Hence, this appendix focuses on issues that were relevant to editing of the 2023 data. Reports that document edits in prior years are available on the [NSDUH Data Collection](#) web page.

For questions or sections of the interview that have changed across years, the issues that are shown in the appendix tables reflect cumulative response patterns that have been observed in the data. However, not all of these patterns may be encountered in a given survey year. In addition, no editing was done for some types of responses that are presented in the tables. Nevertheless, this information is included to document the types of noteworthy responses for which the data were not edited.

B.2 Detailed Documentation of Edits for Selected Back-End Demographics Variables

This section presents tables describing edits for the current education and the employment and workplace variables from the back-end demographics.

- [Table B.1](#) presents edits pertaining to variables in the current education section.
- [Table B.2](#) presents edits pertaining to variables in the employment and workplace section.

¹ All web-based interview questions were self-administered.

Table B.1 Edits Pertaining to the Current Education Section

Issue	Edits Implemented
Respondents (Rs) did not know or refused to report whether they were going to school, reported being on a holiday or break from school (QD17A = 1), but reported that they did not intend to return to school once their break was over (QD17B = 2).	Rs were defined as not going to school (EDUSCHLGO = 2). The variables pertaining to the current grade through the number of days that Rs skipped school in the past 30 days (EDUSCHGRD, EDUFULPAR, EDUSCKMON, EDUSCKEST, EDUSKPMON, and EDUSKPEST) were assigned legitimate skip codes.
For Rs aged 12 to 17 who completed the 11th grade or lower, the current grade (QD18) was potentially inconsistent with the highest grade that the R reported completing (QD11) despite the presence of a consistency check between QD11 and QD18. Also, the consistency check generally applied to the completion of levels of education beyond the 11th grade and therefore was less applicable to Rs aged 12 to 17.	An algorithm was developed for Rs aged 12 to 17 who had completed the 11th grade or lower and who were currently in the 12th grade or lower. The algorithm compared the self-reported current and highest grades with the R's current age (Section 5.2.6). An edited variable for the highest grade completed (EDUHGHEDT) also was created. Edits were generally implemented as follows: ▪ When both the current grade and the highest completed grade were potentially consistent with the R's age, the edits picked the response from QD18 or QD11 that would yield the most consistent data. The second variable in the pair then was edited for consistency with the response that was picked as being most consistent. ▪ When the current grade was more consistent with the R's current age than was the reported highest level of education from the interviewer-administered demographics, EDUHGHEDT was logically assigned a code to indicate that the R had completed the lower grade that was adjacent to the R's current grade. ▪ When the highest completed level of education was more consistent with the R's current age than was the reported current grade, the edited current grade (EDUSCHGRD) was logically assigned a code to indicate that the R was in the next highest grade relative to the one the R had completed, or else EDUSCHGRD was coded as bad data. When neither the current grade (QD18) nor the highest completed grade (QD11) was consistent with the R's age, either EDUHGHEDT or EDUSCHGRD (or both) were coded as bad data. ▪ If the current grade was lower than the expected grade, then the answer that was most consistent with the R's current age was chosen and the other variable in the pair was coded as bad data. ▪ If the current grade was exactly two grades higher than the last grade but the highest grade was lower than the expected highest grade, then EDUHGHEDT was coded as bad data. ▪ If the current grade was more than two grades higher than the last grade but the current grade was lower than the expected current grade, then EDUSCHGRD was coded as bad data. ▪ Otherwise, if the current grade was more than two grades higher than the last grade, then both EDUHGHEDT and EDUSCHGRD were coded as bad data.

Table B.1 Edits Pertaining to the Current Education Section (continued)

Issue	Edits Implemented
Rs were aged 12 to 17 and reported being in college.	No editing was done in the following situations: ▪ The report of currently being in college was consistent with the R's age and highest level of education. ▪ The R was aged 16 or 17, received a high school diploma or a general educational development (GED) certificate, and reported currently being in the first year of college. EDUHGHEDT or EDUSCHGRD (or both) were coded as bad data in the following situations: ▪ If the highest level of education was consistent with the R's current age but the current year in college was not, then EDUSCHGRD was coded as bad data. ▪ If the R was aged 16 or 17 and reported completing the 12th grade without receiving a diploma, then EDUSCHGRD was coded as bad data. ▪ If both the highest level of education and the report of currently being in college were inconsistent with the R's current age, then both EDUHGHEDT and EDUSCHGRD were coded as bad data.
Rs were aged 12 to 17, were in the 1st through 12th grades, and reported receiving a doctorate or professional degree.	EDUHGHEDT was coded as bad data.
Rs were aged 18 or older, were currently in school, completed the 11th grade or lower, were currently in the 12th grade or lower, and the current grade in QD18 was zero, one, or two grade levels higher than the highest grade in QD11.	No editing was done to EDUHGHEDT or EDUSCHGRD in these situations.
Rs were aged 18 or older, were currently in school, completed the 11th grade or lower, were currently in the 12th grade or lower, and the current grade in QD18 was lower than the highest grade or more than two grade levels higher than the highest grade from QD11.	EDUSCHGRD was coded as bad data. No editing was done to EDUHGHEDT.
Rs were aged 18 or older, reported completing the 12th grade without receiving a diploma (QD11 = 13), and were currently in the 12th grade.	No editing was done to EDUHGHEDT or EDUSCHGRD. Adults who had not gotten a high school diploma could report that they were in the 12th grade if they were getting their GED certificate of high school completion.

Table B.1 Edits Pertaining to the Current Education Section (continued)

Issue	Edits Implemented
Rs were aged 18 or older and reported being in college. However, Rs also reported completing the 11th grade or lower or completing the 12th grade without receiving a high school diploma.	The following edits were implemented if Rs were (1) aged 18 or 19 and reported being in the first or second year of college, (2) aged 20 to 22 and reported being in the first through fourth year of college, or (3) aged 23 or older and in the fifth or higher year of college: ▪ If the highest level of education was answered as 1 or 2, then EDUHGHEDT was set to a value of 32 (Regular high school diploma LOGICALLY ASSIGNED) because answers of 1 or 2 in QD11 could be typographical errors for answers of 12. ▪ If the highest level of education was answered as 4, then EDUHGHEDT was set to a value of 34 (GED certificate of high school completion LOGICALLY ASSIGNED) because QD11 = 4 could be a typographical error for an answer of 14. ▪ If the highest grade was the 10th grade or lower, then EDUHGHEDT was coded as bad data because the current year in college was more consistent with the R's age. EDUSCHGRD was coded as bad data in the following situations: ▪ The R had completed the 10th or 11th grades or had completed the 12th grade without receiving a high school diploma. ▪ The R reported receiving a bachelor's degree and was in the 12th grade or lower.
Rs were aged 23 or older, the highest level of education was a high school diploma or lower, and the current grade was lower than the highest level of education.	EDUSCHGRD was coded as bad data.
Rs were aged 23 or older, and the current grade was beyond the 12th grade. The values for the highest level of education and the current grade differed by 2 or more, and the value of the highest level of education plus a factor of 10 was lower than the current year in school.	EDUSCHGRD was coded as bad data.
The highest level of education was a high school diploma, and Rs reported currently being in the 12th grade (regardless of age).	EDUSCHGRD was coded as bad data.
Rs received a GED and reported being in the 1st through 12th grades (regardless of age).	The following age-specific edits were implemented: ▪ EDUHGHEDT was coded as bad data for Rs aged 12 to 17. ▪ EDUSCHGRD was coded as bad data for Rs aged 18 or older.
Rs received some college credit but no degree, and Rs were currently in the 8th grade or lower (regardless of age).	EDUSCHGRD was coded as bad data.
Rs received an associate's or master's degree, and Rs were currently in the 12th grade or lower.	The following age-specific edits were implemented: ▪ EDUHGHEDT was coded as bad data for Rs aged 12 to 17. ▪ EDUSCHGRD was coded as bad data for Rs aged 18 or older.

Table B.1 Edits Pertaining to the Current Education Section (continued)

Issue	Edits Implemented
Rs did not report being full-time students in QD19.	EDUSCKMON, EDUSCKEST, EDUSKPMON, and EDUSKPEST were assigned legitimate skip codes if Rs reported being part-time students. These variables retained values of 98 (blank) if QD19 was answered as “don’t know” (DK) or “refused” (REF). The composite variables EDUSCKCOM (from QD20 and QD20DKRE) and EDUSKPCOM (from QD21 and QD21DKRE) also were assigned legitimate skip codes if Rs reported being part-time students.
Question QD20DKRE was skipped because Rs gave a valid response to QD20.	EDUSCKEST was assigned a legitimate skip code. The composite variable EDUSCKCOM was assigned the value from EDUSCKMON (edited variable corresponding to QD20).
Rs were asked question QD20DKRE because they answered QD20 as DK or REF.	The following edits were implemented: - EDUSCKEST was assigned the response from QD20DKRE. For example, if QD20DKRE also was answered as DK, then EDUSCKEST was coded as 94 (DK). - Responses from EDUSCKEST were used to assign a specific number of days to the composite variable EDUSCKCOM. - If EDUSCKEST had values from 1 to 3 that indicated an exact number of days that the R missed school, then EDUSCKCOM was assigned that number of days. For example, if EDUSCKEST = 1 (absent on 0 days), then EDUSCKCOM was coded as 0. - If EDUSCKEST had a value of 3 or 4 that indicated a range of days, then EDUSCKCOM was assigned the value at the midpoint of the range. For example, if EDUSCKEST = 5 (absent on 6 or more days), then EDUSCKCOM was coded as 18 (i.e., the midpoint of 6 to 30). - If EDUSCKEST = 6 (school not in session), then EDUSCKCOM was coded as 90 (school not in session) to correspond to the response for school not being in session that Rs were instructed to key in QD20. - If EDUSCKEST was coded as DK or REF, then EDUSCKCOM was assigned the code from EDUSCKEST.
Question QD21 was skipped because Rs reported in QD20 that school was not in session in the past 30 days (QD20 = 90).	EDUSKPMON and EDUSKPEST were assigned a legitimate skip code. The composite variable EDUSKPCOM also was assigned a legitimate skip code. EDUSCKMON retained a code of 90.
Question QD21 was skipped because Rs answered QD20 as DK or REF and reported in QD20DKRE that school was not in session in the past 30 days (QD20DKRE = 6).	The composite variable EDUSCKCOM was assigned a code of 90 (school not in session). EDUSKPMON and EDUSKPEST (corresponding to QD21 and QD21DKRE, respectively) and the composite variable EDUSKPCOM were assigned legitimate skip codes.
Question QD21DKRE was skipped because Rs gave a valid response to QD21.	EDUSKPEST was assigned a legitimate skip code. The composite variable EDUSKPCOM was assigned the value from EDUSKPMON (edited variable for QD21).

Table B.1 Edits Pertaining to the Current Education Section (continued)

Issue	Edits Implemented
Rs were asked question QD21DKRE because they answered QD21 as DK or REF.	The following edits were implemented: ▪ EDUSKPEST was assigned the response from QD21DKRE. For example, if QD21DKRE also was answered as DK, then EDUSKPEST was coded as 94 (DK). ▪ Responses from EDUSKPEST were used to assign a specific number of days to the composite variable EDUSKPCOM. – If EDUSKPEST had values from 1 to 3 that indicated an exact number of days that Rs skipped school, then EDUSKPCOM was assigned that number of days. For example, if EDUSKPEST = 1 (skipped school on 0 days), then EDUSKPCOM was coded as 0. – If EDUSKPEST had a value of 3 or 4 that indicated a range of days, then EDUSKPCOM was assigned the value at the midpoint of the range. For example, if EDUSKPEST = 5 (skipped school on 6 or more days), then EDUSKPCOM was coded as 18 (i.e., the midpoint of 6 to 30). – If EDUSKPEST was coded as DK or REF, then EDUSKPCOM was assigned the code from EDUSKPEST.
The number of days in the past 30 that Rs missed school because they were sick and the number of days that they skipped school added up to more than 30.	No editing was done in this situation.
Rs reported being currently enrolled in school but reported skipping school all 30 days in the past 30 days.	A code of 11 was assigned to the school variable EDUSCHLGO. This code was intended to indicate to analysts that there was some uncertainty about whether Rs were currently going to school.

Table B.2 Edits Pertaining to the Employment and Workplace Section

Issue	Edits Implemented
Respondents (Rs) did not know or refused to report in question QD26 whether they worked in the past week. However, Rs also reported in question QD33 (edited variable WRKDPSTYR) that they did not have a job in the past 12 months.	Rs were logically inferred not to have worked in the past week (WRKDPSTWK = 4) and not to have had a job in the past week (WRKHADJOB = 4), where 4 = No LOGICALLY ASSIGNED. Subsequent employment and workplace variables that could be assigned legitimate skip codes were edited as though QD26 and QD27 had been answered as “no.”
Rs reported working in the past week in question QD26. However, they subsequently reported missing work for all 30 of the past 30 days because they were sick (in QD40), did not want to be at work (in QD41), or both. Because the past week was included in the 30 days prior to the interview, it would be inconsistent for Rs to report working in the past week but missing work for every day in the past month.	The following edits were implemented in this situation: ▪ If Rs reported that they missed work for all 30 days in the past month because they were sick, the edited variable (WRKSICKMO) was assigned a bad data code. ▪ If Rs reported missing work for all 30 days in the past month because they did not want to be there, the edited variable (WRKSKIPMO) was assigned a bad data code.
Rs reported working in the past week in question QD26. However, Rs subsequently reported being without a job at some point in the past 12 months in QD37 and reported being without a job during all 52 weeks in the past 12 months in QD38. Because all 52 weeks of the 12-month period prior to the interview would include the week prior to the interview, it would be inconsistent for Rs to report working in the past week but not working for all 52 weeks in the past year.	The edited variable pertaining to the number of weeks without a job in the past 12 months (WRKNJBWKS) was assigned a bad data code.
The reported year when Rs last worked for pay in question QD39A was fewer than 5 years from their birth year (including situations where the year Rs reported last working for pay was earlier than the year they were born).	The edited variables WRKLASTMO and WRKLASTYR were assigned bad data codes.
Rs were not asked question QD34 for whether they were self-employed in the past 12 months when they previously reported not working in the past week because they were self-employed and did not have any business (QD30 = 5).	The edited variable pertaining to self-employment in the past 12 months (WRKSELFEM) was assigned a code to indicate that “yes” could be logically inferred. This edit was done instead of assigning a legitimate skip code.

B.3 Detailed Documentation of Edits for Substance Use Variables

This section presents tables that detail the edits that were applied to variables in the substance use sections.

- [Tables B.3](#) to [B.5](#) present edits pertaining to variables for the frequency of use of different substances in the past 12 months or past 30 days, corresponding to the discussion of procedures in Section 6.2.4.
- [Tables B.6](#) to [B.8](#) present edits pertaining to variables for the initiation of use of a given substance, corresponding to the discussion of editing procedures in Sections 6.2.5 and 6.3.6.

- [Table B.9](#) presents miscellaneous edits pertaining to the tobacco variables.
- [Table B.10](#) presents edits pertaining to how respondents obtained the last prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives) the last time they misused them in the past 12 months.

Table B.3 Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use

Issue	Edits Implemented
The number of days that respondents (Rs) used in the past 12 months was greater than 365 days (e.g., if Rs reported using 31 days per month on average, then $31 \times 12 = 372$ days).	Nonmissing values greater than 365 were trimmed back to a maximum of 365 days. A flag was set to indicate that this trimming was done.
The number of days that Rs reported using in the past 30 days implied less use in the past 12 months than what they reported in the 12-month frequency. For example, if Rs reported using on only 1 day in the past 30 days and did not use on 29 days, the maximum number of days that they could have used in the past 12 months would be 336 days.	The 12-month frequency was trimmed to be consistent with the number of days that Rs reported using in the past 30 days. The relevant flag variable indicated that this trimming was done.
Rs who answered the 30-day frequency questions as “don’t know” (DK) or “refused” (REF) had the opportunity to give their best estimate of the number of days they used in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). However, the maximum possible value for the estimated number of days that Rs reported using in the past 30 days implied less use in the past 12 months than what they reported in the 12-month frequency. For example, if the R reported using a substance on “10 to 19 days” in the past 30 days (i.e., used on a maximum of 19 days in that period), the R would <i>not</i> have used the substance on at least 11 of the remaining past 30 days. It would therefore be inconsistent for the R to have reported using the substance on more than 354 days in the past 12 months.	The 12-month frequency was trimmed to be consistent with the maximum number of days that Rs estimated using in the past 30 days. For example, if the R estimated using a substance on “10 to 19” days in the past 30 days, it was assumed that the R could have used on 354 days in the past 12 months, which would be consistent with use on 19 days in the past 30 days (and nonuse on 11 days). This value would not be considerably different from that derived by assuming use on the minimum number of estimated days in the past 30 days. For example, if the R reported using on 10 to 19 days and used on exactly 10 days, then the R would not have used on 20 days in the past 30 days. In this example, the R could have used on 345 days in the past 12 months.
For Rs who estimated their 30-day frequency, the final 12-month frequency was consistent with some, but not all, values in the range for the estimated 30-day frequency.	No editing was done to the 12-month frequency. For example, if the R reported using a substance on “6 to 9 days” in the past 30 days and reported using on 8 days in the past 12 months, the 12-month frequency was within the range of 6 to 8 days. Thus, a 12-month frequency of 8 days would be consistent with the estimated 30-day frequency, as long as the R used on 6 to 8 days in the past 30 days (and did not use on 9 days in that period). In this situation, minimum and maximum possible values were created for use in statistical imputation to assign a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example, the allowable range for assigning the final 30-day frequency was 6 to 8 days rather than 6 to 9 days.

Table B.3 Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
Rs initiated use at some point in the past 12 months. Relative to when Rs were interviewed, they therefore could not have used over the entire 12-month period. As a consequence, the 12-month frequency was too high relative to the period of time in which Rs used.	The value of the 12-month frequency was trimmed according to the maximum possible allowable number of days that Rs could have used in the past 12 months, relative to when they were interviewed. If Rs answered the 12-month frequency in terms of a total number of days used in that period and the answer was greater than the maximum number of days that they could have used, the 12-month frequency was reduced to this upper limit. The maximum possible allowable number of days was determined from the interview date and the month- and year-of-first use or the date of the R's last birthday, whichever yielded the smallest number of days that the R could have used. If the R reported first using in the survey year but did not report the month when the R first used, it was assumed that the R potentially started use in January of the survey year.
Rs initiated use in the past 12 months but more than 30 days prior to their interview date. Rs also reported use on only 1 day in the past 12 months. However, the recency had a value of 1 (Used in the past 30 days) or 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED).	The 12-month frequency of 1 day was set to bad data. If Rs last used in the past 30 days and they initiated use more than 30 days ago but within the past 12 months, Rs logically would have used on at least 1 day in the past 30 days and on at least 1 other day in the period that was more than 30 days ago but within the past 12 months. Similarly, a recency of 8 indicated that there was uncertainty about whether Rs last used in the past 30 days or more than 30 days ago but within the past 12 months. Consequently, use on only 1 day in the past 12 months and initiation more than 30 days ago within the past 12 months would have forced the recency to be more than 30 days ago. In addition, when the recency had indicated use in the past 30 days and the above edit had changed a 12-month frequency from 1 day to bad data, whatever values that existed in the 30-day variables were retained. In particular, if the 30-day frequency had a missing value, that missing value was retained. (Other edits described below would have logically inferred that Rs used on 1 day in the past 30 days if they also had indicated use on exactly 1 day in the past 12 months.)
Information was needed for establishing constraints on donors for the imputation of the 12-month frequency-of-use variables. For example, if a 12-month frequency variable had been changed from 1 day to bad data, the statistical imputation procedures constrained the imputation of the 12-month frequency to a value greater than 1 day.	Indicator flags were created to aid in establishing specific constraints for the imputation of the 12-month frequency-of-use variables. The indicator flags had the following values: 1 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN 1 DAY. 2 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF USE. 3 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF BINGE ALCOHOL USE. (Applied only to alcohol).

Table B.3 Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
Rs initiated use at some point in the past 30 days. Relative to when Rs were interviewed, however, the 30-day frequency (or the minimum value of the "estimated" 30-day frequency) exceeded the maximum possible number of days that Rs could have used the substance in that period.	If Rs gave an exact number of days for the 30-day frequency, the value was reduced to agree with the maximum possible number of days that they could have used the substance. If Rs estimated the 30-day frequency as a range of a possible number of days, the range for the estimated 30-day frequency was reduced to agree with the maximum possible number of days that they could have used. For example, if the R estimated using a substance on 20 to 29 days in the past 30 days but could have used the substance at most on 15 days in that period, the estimated 30-day frequency was revised to the range indicating use on 10 to 19 days.
The 30-day frequency was greater than the 12-month frequency.	The value for the 30-day frequency was retained, and the 12-month frequency was set to bad data. (Any of the corresponding 12-month frequency source variables that Rs answered were set to bad data as well.) This edit conserved data reported by Rs that suggested more frequent use in the past 30 days. Although both the 12-month and 30-day frequencies could have been set to missing (i.e., bad data) and subsequently imputed, a drawback of this approach is that a donor could be chosen whose 30-day and 12-month frequency data were consistent but whose data were considerably different from what the recipient R originally reported, especially if the recipient R's original answer to the 30-day frequency was at or close to 30.
Rs gave their best estimate of the number of days they used in the past 30 days, but the bottom end of the range for an "estimated" 30-day frequency exceeded the value for the 12-month frequency (i.e., after any opportunities that Rs had for inconsistency resolution).	The value indicating the range for the "estimated" 30-day frequency was retained, and the 12-month frequency was set to bad data. For example, if the R reported using a substance on 5 days in the past 12 months but estimated use of that substance on 6 to 9 days in the past 30 days, there would be no way for these two answers to be consistent. Setting the 12-month frequency to bad data in this situation is consistent with the edit when Rs report an exact number of days for the final 30-day frequency that is greater than the final 12-month frequency.
The bottom end of the range for the final "estimated" number of days that Rs used a substance in the past 30 days was exactly equal to the final 12-month frequency.	The final, edited 30-day frequency was set to the value for the 12-month frequency. The code was retained for the "estimated" 30-day frequency to indicate to analysts where the value came from for the 30-day frequency. For example, if Rs reported using a substance on exactly 6 days in the past 12 months and estimated using that substance on 6 to 9 days in the past 30 days, the 30-day frequency was set to exactly 6 days. The two answers in this example are not necessarily inconsistent if Rs used on exactly 6 days in the past month. Thus, the response of 6 days in the past 12 months could be thought of as narrowing down the possible number of days that Rs could have used in the past 30 days.

Table B.3 Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
Rs reported last using a substance in the past 30 days, and the 30-day frequency had a final response of DK or REF (i.e., Rs did not give a best estimate for the 30-day frequency). However, the 12-month frequency indicated use on exactly 1 day.	The final, edited 30-day frequency was logically inferred to be 1 day. If the R last used in the past 30 days and used on only 1 day in the past 12 months, that 1 day of use had to have occurred in the 30 days prior to the interview.
The value for the final 12-month frequency was somewhere within the range for the final “estimated” number of days that Rs used a substance in the past 30 days.	No further editing was done because the data were not necessarily inconsistent. For example, if the R reported using a substance on 8 days in the past 12 months and estimated use of that substance on 6 to 9 days in the past 30 days, use of the substance on 6, 7, or 8 days in the past 30 days would still be consistent with use on 8 days in the past 12 months. In this situation, minimum and maximum possible values were created for use in statistical imputation to assign a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example where the R reported using on 8 days in the past 12 months and 6 to 9 days in the past 30 days, the allowable range for the 30-day frequency would be 6 to 8 days, not 6 to 9, in order for the 30-day frequency to be consistent with the 12-month frequency.
The overall 12-month frequency did not agree with the reported preference for giving the answer (e.g., if Rs originally indicated that they would give an answer in terms of a total number of days per year, answered the days per year question as DK or REF, and then gave an answer in average days per month). This pattern was a function of the computer-assisted interviewing (CAI) logic, which routed Rs through the 12-month frequency series until they gave a nonmissing answer or answered all 12-month frequency questions as DK or REF. Similarly, Rs could change their preferred way of reporting the 12-month frequency in response to a consistency check with the 30-day frequency (e.g., starting out as reporting in terms of the total number of days in the past 12 months but changing to average number of days per month in response to a consistency check).	The preferred way of reporting the 12-month frequency that best agreed with the final answer to the 12-month frequency was assigned to the “best way” variable (e.g., ALBSTWAY for alcohol), pertaining to Rs’ preferred way of answering the 12-month frequency questions. Thus, if Rs started out as preferring to give an answer in total days per year but the final answer came from an average number of days per month, the preferred method of reporting would be consistent with the reporting in days per month. Any other 12-month frequency questions that Rs answered that did not correspond with the final result were overwritten with LEGITIMATE SKIP Logically assigned codes. In the above example, the initial answer in terms of the total number of days used in the past 12 months would be overwritten, and the final answer of average number of days per month would be retained.
For inhalants and methamphetamine, Rs initiated use more than 30 days ago but within the past 12 months. However, the 12-month frequency and 30-day frequency were equal. Logically, if Rs initiated use more than 30 days ago but within the past 12 months, then the 12-month frequency should be greater than the 30-day frequency.	The 12-month frequency was set to bad data and subsequently imputed, with a lower bound equal to the 30-day frequency plus 1 so that the imputed 12-month frequency would be greater than the 30-day frequency.

Table B.3 Edits Pertaining to 12-Month and 30-Day Frequency Variables for Alcohol through Methamphetamine Other Than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
For inhalants and methamphetamine, Rs who initiated use in the past 30 days had a 12-month frequency that was greater than their 30-day frequency. Logically, if Rs initiated use in the past 30 days, then all use in the past 12 months occurred in the past 30 days.	The 12-month frequency was trimmed to equal the 30-day frequency.
For inhalants and methamphetamine, Rs who were past year initiates and past month users reported a 12-month frequency that was too large relative to when they initiated use in the past year and their 30-day frequency. For example, if an R initiated use more than 30 days ago but within the past 2 months, reported use on 40 days in the past 12 months, but reported use on only 1 day in the past 30 days, then the R could not have used on 39 days in the period more than 30 days ago but within the past 12 months.	The 12-month frequency was trimmed to equal the maximum allowable value that would be consistent with the 30-day frequency.

Table B.4 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine and Crack Cocaine

Issue	Edits Implemented
The 30-day or 12-month frequency of crack use was greater than the corresponding frequency of any cocaine use.	The higher value from the crack frequency was assigned to the cocaine frequency. For the 12-month frequency, this edit also involved moving over the source variables from crack that were associated with the overall crack frequency. In addition, the cocaine variable associated with the preferred method of reporting the 12-month frequency (CCBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the crack cocaine variables.
Respondents (Rs) indicated that their month of first use for any cocaine was in the same month they were interviewed, and they used cocaine on only 1 day in the past 12 months. However, Rs also were users of crack cocaine, and they had missing data for the corresponding 12-month frequency variables pertaining to crack cocaine.	Logically, these Rs also would have used crack cocaine, on 1 day in the past 12 months. Therefore, the corresponding 12-month frequency variable for crack cocaine was set to 1 day in this situation.
Rs were skipped out of the 12-month or 30-day frequency questions for cocaine because of their original answer to the cocaine recency, but nonmissing values existed in the corresponding frequency variables for crack. For example, if Rs reported last using any cocaine more than 30 days ago but within the past 12 months, but reported last using crack within the past 30 days, 30-day frequency data could exist for crack but not for any cocaine.	No editing was done to the 12-month or 30-day cocaine frequency that had been skipped. In the imputation stage, the cocaine 12-month or 30-day frequency was subsequently edited or imputed for consistency with the corresponding crack frequency data.
Rs estimated the frequency of use of both cocaine and crack in the past 30 days, with the estimated frequency for crack being greater than the estimated frequency for any cocaine.	The “estimated” value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that Rs used any cocaine in that period. The code assigned to the “estimated” cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST.
The “estimated” 30-day frequency of crack use was greater than the total number of days that Rs reported using any cocaine in that same period.	The “estimated” value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the “estimated” cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST. The edited cocaine 30-day frequency variable COCUS30A was assigned a bad data code to wipe out the R’s original answer. For example, if the R reported using any cocaine on exactly 5 days in the past 30 days but reported using crack on 6 to 9 days in that period (i.e., a code of 3 in question CK06DKRE), the R was logically inferred to have used any cocaine on 6 to 9 days in the past 30 days. A code of 13 was assigned to CC30EST, corresponding to the code of 3 in CK30EST. In this example, it would be impossible for the R to have used crack on at least 6 days in the past 30 days but to have used any cocaine on only 5 days in that same period.

Table B.4 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine and Crack Cocaine (continued)

Issue	Edits Implemented
The value for the cocaine 30-day frequency was somewhere within the range given by the "estimated" crack 30-day frequency.	All cocaine use in the past 30 days was not automatically inferred to be crack use. Therefore, if the value for the cocaine 30-day frequency was the minimum possible number of days that Rs could have used crack in that period, the crack 30-day frequency CRKUS30A was set to be equal to the cocaine 30-day frequency COCUS30A. For example, if the R reported using cocaine on exactly 6 days in the past 30 days and estimated using crack on 6 to 9 days in that period, it could be logically inferred that the R used crack on exactly 6 days. If the value for the cocaine 30-day frequency was greater than the minimum value for the estimated crack frequency, no further editing was done. However, information regarding the narrower range for the 30-day frequency of crack use was provided for use in statistical imputation of the final value for the crack 30-day frequency. For example, if the R reported using cocaine on 7 days and crack on 6 to 9 days, then the allowable range for the final crack 30-day frequency would be 6 or 7 days.

Table B.5 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use

Issue	Edits Implemented
The frequency of binge alcohol use (i.e., consumption of four or more drinks for females or five or more drinks for males) in the past 30 days (ALCBNG30D, from AL08) was greater than the overall frequency of consumption of any alcohol in the past 30 days (ALCDAYS, from AL06), despite the presence of a consistency check.	If respondents (Rs) reported binge alcohol use on all 30 days, then the edits inferred that they drank on all 30 days. If Rs reported binge alcohol use on fewer than 30 days, then the value from the binge alcohol frequency ALCBNG30D was retained and the overall 30-day frequency ALCDAYS was set to bad data. This approach conserved data reported by Rs that could indicate heavy alcohol use (i.e., consumption of four or more drinks for females or five or more drinks for males on 5 or more days in the past 30 days).
The frequency of binge alcohol use in the past 30 days (ALCBNG30D) was greater than the maximum possible value for the estimated frequency of consumption of alcohol in the past 30 days (AL30EST, from AL06DKRE). For example, if Rs reported binge alcohol use on 10 days in the past 30 days but estimated drinking alcohol on 6 to 9 days in that period, it would be impossible for these two answers to be consistent.	If Rs reported binge alcohol use on all 30 days, then the edits inferred that they drank on all 30 days (i.e., the variable ALCDAYS was assigned a value of 30). This edit also was done if Rs reported having five or more drinks per occasion on all 30 days and they continued to answer the follow-up question AL06DKRE as “don’t know” (DK) or “refused” (REF). If Rs reported binge alcohol use on fewer than 30 days, then the value from ALCBNG30D was retained, and the estimated 30-day frequency variable AL30EST was assigned a bad data value.
Rs drank on 1 day in the past month and the usual number of drinks was four or more (for females) or five or more (for males) on this 1 day, but the frequency of binge alcohol use was 0 (or was answered as DK or REF).	A value of 1 day was assigned to ALCBNG30D. If Rs met the threshold for binge alcohol use on the 1 day when they drank, then they logically had to engage in binge alcohol use on exactly 1 day in the past month.
Rs drank on 1 day in the past month and the usual number of drinks was fewer than four (for females) or fewer than five (for males), but the frequency of binge alcohol use was answered as DK or REF.	A code of 80 was assigned to ALCBNG30D, where 80 = NO OCCASION OF 4+/5+ DRINKS IN PAST 30 DAYS Logically assigned. If Rs reported having a number of drinks on the 1 day when they drank that was below the threshold for binge alcohol use, then these Rs logically had no occasions of binge alcohol use in the past month.
The frequency of binge alcohol use in the past 30 days (ALCBNG30D) fell within the range of the estimated frequency of consumption of alcohol in the past 30 days (AL30EST).	No further editing was done because the data were not necessarily inconsistent. For example, if Rs reported drinking any alcohol on 6 to 9 days in the past 30 days and reported binge alcohol use on 7 days in that period, these data would be consistent as long as Rs drank any alcohol on 7, 8, or 9 days in the past 30 days. In this situation, minimum and maximum possible values were created for use in the imputation of a final 30-day frequency for alcohol. This procedure narrowed the allowable range of the alcohol 30-day frequency variable ALCDAYS for consistency with the binge frequency variable ALCBNG30D. The final 30-day frequency for alcohol was picked from within that narrowed range. In the above example where Rs reported binge alcohol use on 7 days in the past 30 days and 6 to 9 days in that period, the allowable range for the 30-day frequency would be 7 to 9 days, not 6 to 9, in order for ALCDAYS to be consistent with ALCBNG30D.

Table B.5 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use (continued)

Issue	Edits Implemented
Rs drank on more than 1 day in the past month, and the usual number of drinks per day exceeded the gender-specific threshold for binge alcohol use (i.e., four or more for females or five or more for males), but they reported binge drinking on 0 days.	Both the usual number of drinks (i.e., ALCUS30D, from AL07) and ALCBNG30D (from AL08) were set to bad data. This approach was consistent with cognitive testing results that suggested Rs sometimes incorrectly answer AL07 by indicating the total number of drinks they had over the entire 30-day period instead of the usual number of drinks they had per day. Consequently, these edits did not automatically assume that these Rs were binge alcohol users.
The number of days that Rs reported binge alcohol use equaled the total number of days that they reported drinking any alcohol (ALCBNG30D = ALCDAYS), but the reported usual number of drinks was below the gender-specific threshold (or was answered as DK or REF).	A code of 975 was assigned to ALCUS30D (i.e., the edited version of AL07) to indicate usual consumption of at least four drinks (for females) or at least five drinks (for males).
Rs answered the binge alcohol frequency question AL08 as DK or REF and drank on more than 1 day in the past month, but the usual number of drinks on those days when they drank in the past 30 days (from AL07) exceeded the gender-specific threshold for binge alcohol use.	No editing was done to ALCBNG30D to replace the missing data. However, a flag was set to restrict the imputation to donors who reported binge alcohol use on at least 1 day in the past 30 days.

Table B.6 Edits Pertaining to Initiation Variables Based on Consistency Checks

Issue	Edits Implemented
Respondents (Rs) indicated in the final verification check (e.g., CGCC22 for cigarettes) that the age at first use (AFU) or age at first misuse (AFM) based on the new month and year of first use (MFU and YFU) or month and year of first misuse (MFM and YFM) was correct (e.g., CGCC22 = 4). (For brevity, the initiation data subsequently refer only to “first use” and to the AFU, YFU, and MFU data but also apply to first misuse for prescription drugs.) The computer-assisted interviewing (CAI) program updated the AFU (such as AGE1STCG for cigarettes) with the value of the age calculated from the MFU and YFU (referred to subsequently as the MYR1ST age, such as MYR1STCG, for cigarettes). However, the new value for the AFU indicated that Rs were more than 1 year younger than their current age at the time they first used the substance (e.g., the R was 16, reported first use of the substance at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used the substance). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained for the AFU (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable (e.g., CIGYFU for cigarettes), and a code of 89 was assigned to the MFU variable (e.g., CIGMFU) (Section 2.4.2). Consistent with procedures in prior years, therefore, the values for the MFU and YFU were statistically imputed.
The final verification check (e.g., CGCC22) was skipped because Rs entered revised data for the MFU and YFU that made them consistent with the AFU.	No editing was done because Rs were considered to have resolved the inconsistency.
The final verification check (e.g., CGCC22) was skipped because the R entered a new MFU that was the same as the R’s birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R’s birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the MFU and YFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 2005, was interviewed in March 2023 (age 17 at the time of the interview), reported first use of a substance at age 17, and initially reported first use in May 2022. First use in May 2022 would have meant that the R’s AFU was 16 because the R’s 17th birthday was not until June 2022. If the R changed the month and year to June 2022, those answers could be consistent with first use at age 17, if the use occurred after the R’s birthday. However, if the R changed the month and year to June 2021, it would never be possible for the R to have first used at age 17 and also to have first used in June 2021. In this latter situation, the MFU and YFU were set to bad data.
Rs entered a new MFU or YFU that differed from what they previously reported. The MYR1ST age based on the revised MFU and YFU still mismatched the AFU, but Rs indicated in the final verification check that this MYR1ST age was correct.	No editing was done in this situation. The CAI program automatically updated the AFU to be consistent with the updated values reported for the MFU and YFU.

Table B.6 Edits Pertaining to Initiation Variables Based on Consistency Checks (continued)

Issue	Edits Implemented
The consistency check was triggered between the AFU and the MYR1ST age that was calculated from the MFU and YFU. However, Rs answered the first consistency check (e.g., CGCC19 for cigarettes) as “don’t know” (DK) or “refused” (REF). Rs then exited the consistency check loop without having resolved the inconsistency.	The AFU was retained, but the MFU and YFU were set to bad data. Consequently, the MFU and YFU needed to be imputed, but the AFU did not.
Rs entered values in the consistency checks for the MFU and YFU that yielded a nonmissing MYR1ST age based on the MFU and YFU. However, they failed to resolve the inconsistency between the AFU and the MYR1ST age. Rs also reported either in the first verification check that the MYR1ST age was not correct (e.g., CGCC19 = 6 for cigarettes) or reported in the second verification check that the AFU was correct (e.g., CGCC20 = 2 for cigarettes).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: ▪ The default edit was to set the MFU and YFU to bad data. ▪ As an exception to the default edit, if the final verification check was answered as DK or REF, the MFU and YFU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs entered values in the consistency checks for the MFU and YFU that again yielded a nonmissing MYR1ST age based on the MFU and YFU. However, they failed to resolve the inconsistency between the AFU and the MYR1ST age. Rs also reported in the second verification check that neither the AFU nor original MYR1ST age was correct (e.g., CGCC20 = 3 for cigarettes).	The following edits were implemented for the AFU, MFU, and YFU: ▪ The default edit was to set the AFU, MFU, and YFU to bad data. ▪ As an exception to the default edit, if the final verification check was answered as DK or REF, the AFU, MFU, and YFU were assigned the code that corresponded to the answer in the final verification check.
Rs triggered the initial consistency check between the AFU, MFU, and YFU and reported that the MYR1ST age that was calculated from the MFU and YFU was correct (e.g., CGCC19 = 4). However, they answered the second consistency check (e.g., CGCC20 for cigarettes) as DK or REF. Consequently, Rs did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for DK or REF were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R’s correct age when the R first used a substance, or whether the MFU and YFU indicated the R’s correct age at initiation. Therefore, the AFU, MFU, and YFU all were set to missing values.

Table B.7 Edits Pertaining to Initiation Variables Other Than for Parent/Child Relationships

Issue	Edits Implemented
The age at first use (AFU) was greater than the respondent's (R's) reported age. This issue applied to all substances except pipe tobacco (which does not have an AFU question) and to the prescription psychotherapeutic drugs (pain relievers, tranquilizers, stimulants, and sedatives). See Section 6.3.6.2 for discussion of the logic for comparing the age at first misuse (AFM) for prescription drugs with the R's age.	The final age was accepted as the standard. For tobacco through methamphetamine, the inconsistent AFU was set to bad data. Any month and year data associated with the AFU also were set to bad data. For prescription psychotherapeutic drugs, Rs could not proceed with the interview unless they made the AFM consistent with their age or else changed the answer for the AFM to "don't know" (DK) or "refused" (REF).
The AFU had a value of DK or REF, including situations where this assignment has been made from the consistency check data (or hard error for the prescription psychotherapeutic drugs).	For substances other than prescription psychotherapeutic drugs: - If the month of first use (MFU) and year of first use (YFU) questions were skipped because Rs answered the AFU as DK or REF to begin with, the DK or REF value from the AFU was propagated onto the skipped MFU and YFU variables. This edit was designed to indicate the reason that the MFU and YFU variables had been skipped. In addition, because Rs may have first used the substance within 1 year of their current age, the MFU and YFU questions may have been relevant to Rs. - If Rs had answered the MFU and YFU questions but the final AFU had a value of DK or REF (i.e., due to a consistency check response), the MFU and YFU data were overwritten with the corresponding DK or REF value from the edited AFU. Retaining the MFU and YFU data in this situation would imply that Rs used within 1 year of their current age. Responses of DK and REF were not "propagated" for the prescription drug initiation variables because of the volume of variables. However, the skipped initiation variables still had a missing value (i.e., 98, for blank).
The MFU and/or YFU questions were answered because the initial answer to an AFU was within 1 year of the R's current age. However, the final, edited AFU was more than 1 year younger than the R's current age based on the consistency checks between the AFU, YFU, and MFU that were described in Table B.6 .	The original answers to the MFU and YFU were overwritten with logically assigned legitimate skip codes (Section 2.4.2). If Rs originally had answered the relevant AFU as being more than 1 year younger than their current age, the CAI program would have skipped Rs past the questions about the MFU and YFU.
The MFU was skipped because Rs answered the YFU as DK or REF.	For substances other than prescription psychotherapeutic drugs, the DK or REF value from the YFU was propagated onto the skipped MFU. That is, if Rs did not know in what year they first used the substance, it was assumed that they would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year. Responses of DK and REF were not "propagated" from the year of first misuse (YFM) to the month of first misuse (YFM) for the prescription drug initiation variables because of the volume of variables. However, the skipped MFM variables still had a missing value (i.e., 98, for blank).
Rs had missing data for the MFU. However, the AFU and YFU logically restricted the MFU to a single possible month.	The MFU was logically inferred to be the only month when Rs could have initiated use.

Table B.8 Edits Pertaining to Parent/Child Initiation Variables

Issue	Edits Implemented
The age at first use (AFU) for a child substance (e.g., LSDAGE for LSD) indicated first use at age 1 or 2, but the AFU for the parent substance (e.g., HALLUCAGE for any hallucinogen) had a missing value. This issue applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack and hallucinogens/LSD/PCP/Ecstasy or Molly.	Ages of 1 or 2 in the AFU variables for child substances (e.g., LSDAGE) were set to bad data. The imputation procedures for AFU variables did not allow respondents (Rs) with AFUs of 1 or 2 to be donors for imputation. Logically, however, an AFU of 1 or 2 for a child substance would preclude the AFU for the parent substance from being imputed to any value other than 1 or 2. For example, if LSDAGE was 1 or 2 and the AFU for any hallucinogen (HALLUCAGE) had a missing value, the only potential imputation donors for HALLUCAGE would be those Rs with HALLUCAGE of 1 or 2. By definition, however, no one could be a donor for imputation. By setting LSDAGE to bad data, the AFUs for both LSD and any hallucinogen would be imputed.
The AFU for a child substance (e.g., LSDAGE) indicated first use at age 3, but the AFU for the parent substance (e.g., HALLUCAGE) had a missing value. This issue applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack and hallucinogens/LSD/PCP/ Ecstasy or Molly.	No editing was done when this pattern occurred. Logically, however, the AFU for the parent substance had to have a value of 3 because Rs with AFUs of 1 or 2 (e.g., HALLUCAGE of 1 or 2) were not eligible to be donors in the imputations. The final imputed AFU (e.g., IRHALLUCAGE for any hallucinogen) was set to a value of 3.
The AFU for a child substance (e.g., crack cocaine) was younger than the AFU for the parent substance (e.g., any cocaine). This issue applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack and hallucinogens/LSD/PCP/Ecstasy or Molly.	This edit required the AFUs to be defined (i.e., not missing) for both the parent and child substances. The younger child AFU was assigned to the parent AFU. For example, if the AFU of crack cocaine was younger than the AFU of any cocaine, then the edits assigned the value from the crack AFU to the cocaine AFU.
Rs in the hallucinogens section were lifetime users only of LSD, only of PCP, or only of Ecstasy or Molly. This issue included situations in which the only reported use of a hallucinogen was “some other hallucinogen,” but only LSD, only PCP, or only Ecstasy or Molly was specified. The CAI logic skipped these Rs out of questions related to the age, month, and year that they first used the specific hallucinogen.	Data on initiation of hallucinogen use were moved over to the corresponding specific hallucinogen variables that had been skipped. For example, if Rs had used only LSD, the hallucinogen age (from HALLUCAGE) was assigned to the LSD AFU (LSDAGE). Similarly, if Rs had provided data on the month and year that they first used any hallucinogen, these data were assigned to the LSD month and year variables LSDMFU and LSDYFU. Similar edits were done when Rs had been skipped out of questions related to specific hallucinogens but were logically inferred to be lifetime users only of LSD, only of PCP, or only of Ecstasy or Molly, based on the “OTHER, Specify” responses.
Rs in the hallucinogens section were lifetime users of two or more of the following: LSD, PCP, or Ecstasy or Molly. No other hallucinogen use was indicated. However, Rs answered the AFU question for any hallucinogen (edited variable HALLUCAGE) as “don’t know” or “refused.” In addition, the specific hallucinogens that Rs had used had legitimate values for their AFU variables (i.e., LSDAGE, PCPAGE, or ECSTMOAGE, for LSD, PCP, or Ecstasy or Molly, respectively).	Logically, the AFU for any hallucinogen (HALLUCAGE) had to be the minimum of the ages indicated in LSDAGE, PCPAGE, or ECSTMOAGE. Therefore, HALLUCAGE was assigned the minimum value from LSDAGE, PCPAGE, or ECSTMOAGE. If this edit was done and month and year data were available for a specific hallucinogen, the month and year variables for any hallucinogen (HALLUCMFU and HALLUCYFU) were edited accordingly. For example, if a missing value in HALLUCAGE was replaced with the value from LSDAGE, and the month of first use (MFU) and year of first use (YFU) for LSD (LSDMFU and LSDYFU) had valid values, the values from LSDMFU and LSDYFU were assigned to HALLUCMFU and HALLUCYFU, respectively. This edit was done only to replace missing data in HALLUCAGE. No editing was done if HALLUCAGE had a value that was lower than the AFU values reported in LSDAGE, PCPAGE, or ECSTMOAGE.

Table B.8 Edits Pertaining to Parent/Child Initiation Variables (continued)

Issue	Edits Implemented
Rs in the cocaine/crack or hallucinogens sections indicated first use of any substance in that category (e.g., any hallucinogen) at their current age but had missing data for the age when they first used specific substances in that category (e.g., LSD).	Rs were logically inferred to have first used the specific substance at their current age.
The AFU for the parent substance was at the R's current age and the child AFU was at 1 year below the R's current age. Consequently, the child AFU had been assigned to the parent AFU.	The associated MFU and YFU from the child data (e.g., for crack cocaine) were assigned to the parent substance's MFU and YFU data (e.g., for any cocaine). However, this edit was not done in the hallucinogens section when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy or Molly. Consequently, some inconsistencies could remain in the edited hallucinogens data.
Rs in the cocaine/crack or hallucinogens sections indicated first use for the parent substance category (e.g., any hallucinogen) in the current survey year but had missing data for the year when they first used a child substance (e.g., LSD).	Rs were logically inferred to have first used the child substance in the current year. If Rs also reported first use of substances in the parent category (e.g., any hallucinogen) in the same month that they were interviewed and they had missing MFU data for child substances within that category, Rs also were logically inferred to have first used the child substance in the same month that they were interviewed. Otherwise, if Rs originally had reported a month when they first used a child substance but that month had been assigned a code of bad data due to the "flag and impute" edits, the MFU (e.g., LSDMFU) also was reset to the originally reported value if it matched the MFU for any substance in the parent category (e.g., HALLUCMFU).
Rs in the cocaine/crack or hallucinogens sections indicated first using any substance in that category (e.g., any hallucinogen) in the year immediately prior to the current one, but that point of initiation was within 12 months of their interview date. These Rs also had missing data for the year and month when they first used child substances in that category (e.g., LSD).	If Rs originally had reported year and month data for when they first used a child substance but these data had been assigned codes for bad data due to the "flag and impute" edits, this information was restored <i>only</i> if the original answers matched the MFU and YFU for any substance in that category. For example, if LSDMFU and LSDYFU had been set to bad data, but the original answers for the LSD month and year matched the month and year in HALLUCMFU and HALLUCYFU, then LSDMFU was equated to HALLUCMFU and LSDYFU was equated to HALLUCYFU.
The MFU and/or YFU questions for the parent substance were answered because the initial answer to the parent AFU was within 1 year of the R's current age. However, the final, edited AFU for the parent substance was more than 1 year younger than the R's current age (e.g., if the unedited AFU for cocaine was within 1 year of the R's current age, but the AFU of crack was more than 1 year younger than the R's current age).	The original answers to the parent MFU and YFU were overwritten with logically assigned legitimate skip codes (Section 2.4.2). If Rs originally had answered the parent AFU as being more than 1 year younger than their current age, the CAI program would have skipped them past the questions about the parent MFU and YFU.
The AFUs for the parent and child substances (or behaviors, in the situation for daily cigarette use) were the same, but the MFU or YFU was earlier for the child substance (e.g., crack) or behavior (for daily cigarette use).	The earlier month and year data from the child substance (e.g., crack) were assigned to the parent substance's month and year data (e.g., cocaine). However, this edit was not done in the hallucinogens section when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy or Molly but the YFU and MFU data for the parent and child substances were inconsistent. Consequently, some inconsistencies could remain in the edited hallucinogens data.

Table B.8 Edits Pertaining to Parent/Child Initiation Variables (continued)

Issue	Edits Implemented
The AFU for any cigarette use (CIGTRY) equaled the age at initiation of daily cigarette use (CIGAGE). However, Rs reported initiating daily cigarette use in a year or month that was earlier than what they reported for first use of any cigarette.	The month and year data for daily cigarette use (CIGDLMFU and CIGDLYFU, respectively) were assigned to the month and year variables for any cigarette use (CIGMFU and CIGYFU, respectively).
For hallucinogens, Rs were known never to have used any substances other than the child substances, the AFU for the parent substance was nonmissing, only one child AFU was missing, and the minimum of the nonmissing child AFUs was greater than the parent AFU.	The missing child AFU variable was assigned to the parent AFU value.
For hallucinogens, Rs were known never to have used any substances other than the child substances, the parent AFU was nonmissing, only one child AFU was missing, the minimum of the nonmissing child AFUs was equal to the parent AFU, and the earliest of the nonmissing child MFUs and/or YFUs was later than the parent MFU and/or YFU.	The missing child AFU variable was assigned to the parent AFU value.
The parent AFU was equal to the R's current age, and some or all of the AFUs for the child substances were missing.	All missing child AFU variables were assigned to the R's current age.
The parent AFU was equal to 1 year less than the R's current age, the child recency ¹ was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child AFU value was missing.	The child AFU variable was assigned to 1 year less than the R's current age.
The AFU for cigarettes was equal to 3 years less than the R's current age, the cigarette recency was lifetime but not past 3 years, and the AFU for daily cigarettes was missing.	The AFU for daily cigarettes was assigned to 3 years less than the R's current age.
Rs reported smoking cigarettes on all 30 days in the past 30 days, but they reported initiating daily cigarette use within 30 days of their interview date. The latter could occur in one of the following ways: (1) the reported month and year for when Rs started smoking cigarettes daily was within 30 days of their interview date; or (2) Rs reported starting to smoke cigarettes every day at their current age (i.e., CIGAGE = AGE), but they had been at this age for fewer than 30 days.	The following edits were implemented: - If only the month and year indicated that Rs had been smoking daily for fewer than 30 days (i.e., but Rs had been at the age given in CIGAGE for at least 30 days), the month and year (CIGDLMFU and CIGDLYFU) were set to bad data. - If Rs had been at the age reported in CIGAGE for fewer than 30 days and reported starting to smoke daily in the same month that they were interviewed, CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data. If Rs had been at the age reported in CIGAGE for fewer than 30 days and CIGDLMFU and CIGDLYFU had codes of "don't know" or "refused," CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data.

Table B.8 Edits Pertaining to Parent/Child Initiation Variables (continued)

Issue	Edits Implemented
Rs reported smoking cigarettes in the past 30 days. Rs did not report smoking on all 30 days in the past 30 days but reported initiating daily cigarette use within 30 days of their interview date.	No editing was done when this pattern occurred, even though question CG15 asked Rs if they ever smoked cigarettes every day for at least 30 days. Consequently, data were preserved that indicated that Rs recently initiated daily cigarette use, even if they could not have smoked daily for at least 30 days.

¹ Because no recency question was associated with daily cigarettes, the overall cigarette recency was used instead.

Table B.9 Miscellaneous Edits Pertaining to Tobacco and Nicotine Vaping Variables

Issue	Edits Implemented
Respondents (Rs) initiated use of cigarettes, smokeless tobacco, cigars, or nicotine vaping at some point in the past 30 days. Relative to when Rs were interviewed, however, the 30-day frequency (or the minimum value of the “estimated” 30-day frequency) exceeded the maximum possible number of days that Rs could have used the substance in that period.	If Rs gave an exact number of days for the 30-day frequency, the value was reduced to agree with the maximum possible number of days that they could have used the substance. If Rs estimated the 30-day frequency as a range of a possible number of days, the range for the estimated 30-day frequency was reduced to agree with the maximum possible number of days that they could have used.
Rs who answered 30-day frequency questions for cigarettes, smokeless tobacco, cigars, or nicotine vaping as DK or REF were asked to give their best estimate of the number of days they used these tobacco products in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). Because there were no 12-month frequency questions for tobacco products or nicotine vaping, inconsistencies would not exist between these “estimated” 30-day frequency variables and other data for these tobacco products or nicotine vaping.	No editing was done to the corresponding edited 30-day frequency variables (i.e., leaving the original answer of DK or REF). A final value that was consistent with the “estimated” 30-day frequency variable was assigned as part of the final imputation process for the 30-day frequency data.
The CAI program had skipped lifetime cigarette questions pertaining to periods where Rs had smoked cigarettes daily for at least 30 days (CG15) or whether they had smoked 100 or more cigarettes (CG16A). Answers to prior questions already indicated that Rs had engaged in these behaviors. For example, if Rs had reported smoking every day in the past 30 days, then they had to have had a point in their life when they smoked cigarettes every day for at least a month. Similarly, it could be inferred from answers to questions about the number of days that Rs smoked in the past 30 days and the number of cigarettes they usually smoked per day that they had smoked at least 100 cigarettes in their lifetime.	It was logically inferred in CIGDLYMO (the edited version of CG15) or in CIG100LF (the edited version of CG16A) that Rs had engaged in the behavior of interest (i.e., ever smoked cigarettes daily for at least 30 days or smoked 100 or more cigarettes in their lifetime). A code of 5 was assigned. This code had the following meaning: 5 = Yes LOGICALLY ASSIGNED (from skip pattern). This code was used instead of a legitimate skip code to signify that the CAI logic had skipped Rs out of the question because they had already provided data to indicate that the question should have been answered affirmatively.
Rs were routed to the lifetime 100-cigarette question (CG16A) because they estimated the number of days that they smoked cigarettes in the past 30 days. Rs did not answer CG16A as “yes,” but the minimum number of days that they could have smoked in the past 30 days and the minimum usual number of cigarettes smoked per day suggested that they smoked 100 or more cigarettes in their lifetime.	It was logically inferred in CIG100LF (i.e., the edited version of CG16A) that Rs had smoked 100 or more cigarettes in their lifetime. For example, if Rs reported smoking 10 to 19 days and reported smoking 16 to 25 cigarettes per day, then they would have smoked at least 100 cigarettes in the past 30 days, even if they smoked on only 10 days and smoked only 16 cigarettes per day.

Table B.10 Edits Pertaining to the Source of Prescription Psychotherapeutic Drug Variables

Issue	Edits Implemented
Respondents (Rs) reported that they got the last drug they misused in “some other way,” but they specified getting it in one of the ways that they had been asked about previously.	The code from the “OTHER, Specify” variable plus a value of 10 was added to the variable for how Rs obtained the drug the last time they misused it (i.e., “bump” by 10). For example, if Rs reported getting the last pain reliever they misused in some other way (PRYGOT = 8), but they specified getting them from a friend or relative for free (i.e., PNRGOTOTH = 4), then PNRGOTLAS was coded as 14. Similar edits were implemented if Rs reported that friends or relatives obtained a drug that Rs last used in “some other way” and then Rs specified one of the ways they had asked about previously. For example, if PNRFFGTSP was coded as 4 (i.e., the friend or relative got the pain reliever from another friend or relative for free), then PNRFFGOT (how the friend or relative obtained the last pain reliever that the R misused) was coded as 14.
Rs reported that they got their last prescription drug through a prescription, either from one doctor or from more than one doctor. However, the only way that Rs reported misusing any prescription drug in that category in the past 12 months was without a prescription of their own (e.g., in question PRYWAYS for pain relievers).	No editing was done to the report of obtaining the last prescription drug that was misused through a doctor’s prescription or to the report of use without a prescription as the only way that Rs misused any prescription drug in that category in the past 12 months. However, a flag was set (e.g., PNRNORXFG) to alert analysts to the occurrence of this pattern.
Rs were logically inferred to have gotten the last medication they misused from friends or relatives for free (e.g., PNRGOTLAS = 14 for pain relievers). Consequently, Rs were not asked how the friends or relatives obtained the medication that Rs last used.	The corresponding edited variable pertaining to how the friends or relatives obtained the medication that Rs last used retained a code of 98 (i.e., blank).
Rs reported that they or friends/relatives got a drug for Rs’ last misuse in “some other way.” However, Rs explicitly specified that the drug they or friends/relatives obtained was an over-the-counter (OTC) drug.	No editing was done when Rs specified that they or their friends or relatives had obtained OTC drugs.

B.4 Detailed Documentation of Edits for Variables in Additional Sections

This section presents tables that detail the edits that were applied to variables in additional questionnaire sections. Because the content of individual sections following the initial substance use sections differs considerably from section to section, many of the edits are specific to a single section of the interview.

- [Tables B.11](#) and [B.12](#) present edits pertaining to variables in the special drugs section, corresponding to the discussion of procedures in Section 7.2.
- [Tables B.13](#) and [B.14](#) present edits pertaining to variables in the blunts section, corresponding to the discussion of procedures in Section 8.2.
- [Tables B.15](#) and [B.16](#) present edits pertaining to variables in the prior substance use section, corresponding to the discussion of procedures in Section 8.3.

- [Table B.17](#) presents edits pertaining to variables in the alcohol and drug treatment section, corresponding to the discussion of procedures in Section 8.4.
- [Tables B.18](#) and [B.19](#) present edits pertaining to variables in the consumption of alcohol section, corresponding to the discussion of procedures in Section 8.5.
- [Table B.20](#) presents edits pertaining to variables in the special topics section, corresponding to the discussion of procedures in Section 9.3.
- [Table B.21](#) presents edits pertaining to editing of variables in the mental health services utilization section, corresponding to the discussion of procedures in Section 10.3.

Table B.11 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Section

Response Pattern	Edit
Variables were skipped because respondents (Rs) never used the substance of interest, and there were no other indications elsewhere in the special drugs section that Rs ever used this substance.	A code of 91 was assigned to the edited variables. For example, if Rs never used heroin, the edited variables HERSMOKE, HRSMKREC, HERSNIFF, HRSNFREC, HERNEEDL, and HRNDLREC were assigned a code of 91.
Variables were skipped because Rs refused to indicate in the corresponding substance use section whether they ever used the substance of interest, and there were no other indications elsewhere in the special drugs section that they ever used this substance.	A code of 97 (i.e., refused) was assigned to the edited variables. Thus, for example, a refusal from the heroin recency-of-use variable was propagated onto the heroin variables in the special drugs section.
Variables were skipped because Rs did not know in the corresponding substance use section whether they ever used the substance of interest, and there were no other indications elsewhere in the special drugs section that they ever used this substance.	The skipped special drugs variables pertaining to this substance retained a value of 98 (i.e., blank).
For cocaine and methamphetamine, there were no indications elsewhere in the special drugs section of Rs reporting use of the substance as “some other drug” that they used with a needle. Rs were lifetime users of the substance of interest, but the corresponding needle recency variable was skipped because of the following: - Rs never used the substance with a needle. - Rs refused to indicate whether they had ever used the substance with a needle. - Rs did not know whether they had ever used that substance with a needle.	- A code of 93 was assigned to the corresponding needle recency variable (e.g., CONDLREC) to indicate that Rs used the substance but never with a needle. - A code of 97 was assigned to the edited needle recency variable (i.e., the refusal was propagated). - The edited needle recency variable retained a code of 98 (i.e., blank).
Rs were lifetime users of heroin, but relevant recency variables for smoking heroin (HRSMKREC), sniffing heroin (HRSNFREC), or using it with a needle (HRNDLREC) were skipped because of the following: - Rs never used heroin via the route of interest. - Rs refused to indicate whether they had ever used heroin via the route of interest. - Rs did not know whether they had ever used heroin via the route of interest. (For heroin use with a needle, there were no other indications elsewhere in the special drugs section of heroin use with a needle.)	- A code of 93 was assigned to the relevant heroin recency variable(s) (e.g., HRSMKREC for smoking heroin) to indicate that Rs used heroin but not in that particular way. - A code of 97 was assigned to the relevant heroin recency variable(s) (i.e., the refusal was propagated). - The edited heroin recency variable(s) retained a code of 98 (i.e., blank). The special situation in which Rs reported lifetime use of heroin in the heroin use section but reported that they never smoked, sniffed, or used it with a needle is discussed in Section 7.2.3.

Table B.11 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Section (continued)

Response Pattern	Edit
General needle use variables were skipped because Rs reported never using cocaine, heroin, methamphetamine, or any other drug (question SD15; edited variable OTDGNEDL) with a needle.	A code of 99 (i.e., legitimate skip) was assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLN, GNNDLSH2, and GNNDGET).
General needle use variables were skipped because question SD15 was answered as “no” (OTDGNEDL = 2). There were no affirmative reports of the use of cocaine, heroin, or methamphetamine with a needle, but one or more of the lifetime needle use variables for these substances was answered as “don’t know” (DK) or “refused” (REF).	When there was no affirmative report of the use of cocaine, heroin, or methamphetamine with a needle, question SD15 was worded as follows: “Have you ever, even once, used a needle to inject <i>any drug</i> that was not prescribed for you?” (wording not italicized in the interview). Therefore, a code of 99 (i.e., legitimate skip) was assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLN, GNNDLSH2, and GNNDGET) because it could be inferred that the response of “no” in question SD15 pertained to use of any drug with a needle. However, no editing was done to any responses of DK or REF in the lifetime needle use variables pertaining to cocaine (COCNEEDL), heroin (HERNEEDL), or methamphetamine (METHNEEDL).
General needle use variables were skipped because question SD15 was refused (OTDGNEDL = 97), and there were no affirmative reports of heroin, methamphetamine, other stimulants, or cocaine with a needle.	The refusal from OTDGNEDL was propagated to the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLN, GNNDLSH2, and GNNDGET.

Table B.12 Edits in the Special Drugs Section (Other Than Skip Patterns) Pertaining to Heroin Use and Needle Use

Issue	Edits Implemented
Respondents (Rs) reported injecting other substances in question SD15, and they specified injecting only cocaine, only heroin, or only methamphetamine in the "OTHER, Specify" questions SD15SP to SD15D. However, Rs reported more recent use of the "other" substance with a needle in question SD16 than was reported in the corresponding needle recency question for cocaine, heroin, or methamphetamine.	The value from SD16 (edited variable OTDGNDLRC) plus a value of 10 (i.e., "bump" by 10) was assigned to the needle recency variable for the specific substance. For example, if the heroin recency indicated that Rs last used heroin with a needle more than 12 months ago (HRNDLREC = 3), they reported the injection of only heroin as some other drug in their lifetime, and the most recent use of "other" drugs with a needle was more than 30 days ago but within the past 12 months (OTDGNDLRC = 2), then HRNDLREC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). This edit was not implemented if Rs had responses of "don't know" (DK) or "refused" (REF) in the "OTHER, Specify" questions SD15SP to SD15D because these responses potentially indicated the injection of substances other than cocaine, heroin, or methamphetamine.
Rs reported injecting other drugs in question SD15, and they specified injecting only cocaine, only heroin, or only methamphetamine in the "OTHER, Specify" questions SD15SP to SD15D. However, Rs reported more recent use with a needle in the corresponding needle recency question for cocaine, heroin, or methamphetamine than they reported in the needle recency question SD16 for other drugs.	The value from the specific needle recency variable (i.e., CONDLREC for cocaine, HRNDLREC for heroin, or METHNDLRC for methamphetamine) plus a value of 10 (i.e., "bump" by 10) was assigned to the needle recency variable OTDGNDLRC (corresponding to question SD16) for other drugs. For example, if the heroin recency indicated that Rs last used heroin with a needle more than 30 days ago but within the past 12 months (HRNDLREC = 2), they reported the injection of only heroin as some other drug in their lifetime, and the most recent use of "other" drugs with a needle was more than 12 months ago (OTDGNDLRC = 3), then OTDGNDLRC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). This edit was not implemented if Rs had responses of DK or REF in the "OTHER, Specify" questions SD15SP to SD15D because these responses potentially indicated the injection of substances other than cocaine, heroin, or methamphetamine.

Table B.12 Edits in the Special Drugs Section (Other Than Skip Patterns) Pertaining to Heroin Use and Needle Use (continued)

Issue	Edits Implemented
Rs were lifetime users of heroin and reported smoking, sniffing, or injecting heroin with a needle at least once in their lifetime, but they did not know or refused to indicate when they last smoked, sniffed, or injected heroin.	The edits typically depended on the most recent use of heroin reported in the corresponding heroin recency variable, except where noted. ▪ In general, the edited heroin recency variables in the special drugs section (HRSMKREC, HRSNFREC, HRNDLREC) were assigned a code of 9 (i.e., used at some point in the lifetime). Exceptions are noted below. ▪ If the only “other” drug that Rs injected in their lifetime in SD15SP to SD15D was heroin and the needle recency for other drugs (OTDGNDLRC, corresponding to question SD16) indicated last use in a definite period, then the value from OTDGNDLRC plus a value of 10 (i.e., “bump” by 10) was assigned to HRNDLREC. In this situation, for example, if OTDGNDLRC = 2 (More than 30 days ago but within the past 12 months), then HRNDLREC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). ▪ If the heroin recency indicated that Rs last used heroin more than 12 months ago and there was no other indication in the special drugs section that they had used heroin more recently, then the edited variables pertaining to smoking, sniffing, or injecting heroin were assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED). This edit did not apply if Rs did not know or refused to report when they last used heroin in a particular way (e.g., smoking it) but reported last using it a different way in the past 12 months (e.g., with a needle).
Rs were lifetime users of cocaine and reported using it with a needle at least once in their lifetime. However, Rs did not know or refused to indicate when they last used cocaine with a needle.	The edits typically depended on the most recent use reported in the corresponding recency variable COCREC, except where noted. ▪ In general, the edited needle recency variable CONDLREC for cocaine was assigned a code of 9 (i.e., used at some point in the lifetime). Exceptions are noted below. ▪ If the only “other” drug that Rs injected in their lifetime in SD15SP to SD15D was cocaine and the needle recency for other drugs (OTDGNDLRC, corresponding to question SD16) indicated last use in a definite period, then the value from OTDGNDLRC plus a value of 10 (i.e., “bump” by 10) was assigned to CONDLREC. In this situation, for example, if OTDGNDLRC = 2 (More than 30 days ago but within the past 12 months), then CONDLREC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). ▪ If COCREC indicated that Rs last used cocaine more than 12 months ago and there was no other information to indicate that they used cocaine with a needle more recently, then CONDLREC was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).

**Table B.12 Edits in the Special Drugs Section (Other Than Skip Patterns)
Pertaining to Heroin Use and Needle Use (continued)**

Issue	Edits Implemented
Rs were lifetime users of methamphetamine and reported using it with a needle at least once in their lifetime. However, Rs did not know or refused to indicate when they last used methamphetamine with a needle.	The edits typically depended on the most recent use reported in the corresponding recency variable METHAMREC, except where noted. ▪ In general, the edited needle recency variable METHNDLRC for methamphetamine was assigned a code of 9 (i.e., used at some point in the lifetime). Exceptions are noted below. ▪ If the only “other” drug that Rs injected in their lifetime in SD15SP to SD15D was methamphetamine and the needle recency for other drugs (OTDGNDLRC, corresponding to question SD16) indicated last use in a definite period, then the value from OTDGNDLRC plus a value of 10 (i.e., “bump” by 10) was assigned to METHNDLRC. In this situation, for example, if OTDGNDLRC = 2 (More than 30 days ago but within the past 12 months), then METHNDLRC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). ▪ If METHAMREC indicated that Rs last used methamphetamine more than 12 months ago and there was no other information to indicate that they used methamphetamine with a needle more recently, then METHNDLRC was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).
Rs reported in the corresponding substance use sections that they never used heroin (HERREC = 91), never used cocaine (COCREC = 91), or never used methamphetamine (METHAMREC = 91). However, Rs specified use of one or more of these substances as “some other drug” that they had ever injected.	No editing was done to the substance use data that appeared earlier in the interview (Section 2.4.1). However, Rs were logically inferred in the special drugs data to be a lifetime user of that substance with a needle, even though the earlier substance use data indicated that Rs never used that substance. If Rs reported the injection of more than one other substance in their lifetime, then the corresponding needle recency variable was set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). For example, if Rs reported in the heroin section that they never used heroin but specified injection of heroin as “some other drug,” the lifetime heroin needle use variable HERNEEDL was set to 3 (Yes LOGICALLY ASSIGNED), and the heroin needle recency HRNDLREC was set to 9. Similar edits were done for the needle use variables pertaining to cocaine and methamphetamine. However, the edits that were described previously applied if Rs reported the injection of only heroin, only cocaine, or only methamphetamine in the “OTHER, Specify” questions SD15SP to SD15D. For example, if HERREC = 91, Rs specified the injection of only heroin in SD15SP to SD15D, and they reported last injecting other drugs with a needle more than 30 days ago but within the past 12 months (OTDGNDLRC = 2), then HERNEEDL was set to 3, as described above. However, HRNDLREC was set to 12 (More than 30 days ago but within the past 12 months LOGICALLY ASSIGNED) instead of being set to 9.

**Table B.12 Edits in the Special Drugs Section (Other Than Skip Patterns)
Pertaining to Heroin Use and Needle Use (continued)**

Issue	Edits Implemented
Rs specified lifetime use of more than five drugs with a needle.	Priority was given to retaining as many unique mentions as possible for other drugs that Rs used with a needle. Thus, multiple mentions of the same drug would be counted only once. Priority also was given to retaining mentions of drugs that were covered in the special drugs section that Rs had not previously reported using with a needle (e.g., if the question corresponding to METHNEEDL had been answered as “no” but methamphetamine had been specified as “some other drug” that Rs used with a needle). Conversely, retention of “OTHER, Specify” codes corresponding to drugs that Rs had already reported using with a needle were given lower priority. If there were still more than five mentions of unique drugs after the above steps, priority was given to retaining the most serious drugs according to the Drug Enforcement Administration (DEA) drug schedule (e.g., first priority to retaining mention of Schedule I drugs that have no approved medical use in the United States, second priority to retaining Schedule II drugs). Finally, after the drugs had been ranked according to their severity based on the DEA drug schedule, if mention of more than five drugs still remained, the codes were retained in the order they appeared in the data.
Rs reported using a needle to inject a drug (SD05 = 1), but they previously reported never using or misusing marijuana, cocaine, heroin, hallucinogens, inhalants, methamphetamine, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives.	No editing was done if Rs - specified needle use involving a drug that could be misused or that had psychoactive properties (e.g., steroids, one or more categories of drugs covered in the earlier substance use sections that were not covered elsewhere in special drugs, such as prescription pain relievers), or - reported one or more “risky” needle use behaviors (reusing a needle, needle sharing, or cleaning a needle with bleach). Rs were inferred not to be lifetime users of needles to inject drugs that were not prescribed for them (OTDGNEDL = 4) if they - specified use of a drug that was typically not misused and did not have psychoactive properties (e.g., if injection only of antibiotics was specified), and - reported never reusing a needle, sharing a needle (before or after someone had used it), or cleaning a needle with bleach (i.e., “risky” needle use behaviors). When OTDGNEDL = 4, any data in the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLN, GNNDLSH2, and GNNDGET were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned).

**Table B.12 Edits in the Special Drugs Section (Other Than Skip Patterns)
Pertaining to Heroin Use and Needle Use (continued)**

Issue	Edits Implemented
The general needle use items corresponding to questions SD12 through SD16 had been answered. However, the only indications of lifetime needle use had been set to bad data because of patterned responses in the corresponding substance use section.	The edited general needle use variables GNNDREUS through GNNDGET were set to bad data.
Rs reported getting their last needle “some other way” and specified a meaningful response for how they last got the needle.	The final, edited variable pertaining to how Rs got their last needle (GNNDGET) was a composite of the response categories that were offered to Rs (i.e., bought the needle from a pharmacy, got the needle from a needle exchange, bought the needle on the street, got the needle in a shooting gallery, got the needle some other way). This creation of a composite variable was done because the computer-assisted interviewing (CAI) logic did not allow Rs to specify an “other” way that they got the needle if they reported getting the needle in one of the first four ways. If Rs reported getting the needle “some other way” and specified a meaningful way they got the needle, then the “OTHER, Specify” response was assigned to GNNDGET.
Rs reported getting their last needle “some other way” and did not know what that other way was, refused to specify what that other way was, or gave a response that was coded as bad data (e.g., a nonsensical response).	The final, edited variable pertaining to how Rs got their last needle (GNNDGET) retained a nonspecific code of “some other way.” Stated another way, the response of “some other way” was given precedence over the missing value in the “OTHER, Specify” response. The edit was done in this manner to provide a nonmissing value for analysts to use.
Rs answered DK or REF at the outset, when asked how they got the last needle that they used (GNNDGET).	The response of DK or REF was retained in the final, edited variable (GNNDGET).

Table B.13 Edits in the Blunts Section Pertaining to Consistency Checks for Initiation

Issue	Edits Implemented
Respondents (Rs) indicated in the final verification check (BLCC06) that the age at first use (AFU) based on the new month and year of first use (MFU and YFU) was correct (i.e., BLCC06 = 4). The computer-assisted interviewing (CAI) program updated the AFU (AGE1STBL) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBL). However, the new value for the AFU indicated that Rs were more than 1 year younger than their current age at the time they first used blunts (e.g., the R was 16, reported first use of blunts at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used blunts). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained in the edited AFU variable BLNTAGE (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable BLNTYFU, and a code of 89 was assigned to the MFU variable BLNTMFU (Section 2.4.2).
The final verification check (BLCC06) was skipped because Rs entered revised data for the MFU and YFU that made BLNTMFU and BLNTYFU consistent with the AFU in BLNTAGE.	No editing was done because Rs were considered to have resolved the inconsistency.
The final verification check (BLCC06) was skipped because Rs entered a new MFU that was the same as their birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, BLNTMFU and BLNTYFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 2005, was interviewed in March 2023 (age 17 at the time of the interview), reported first use of blunts at age 17, and initially reported first use in May 2022. First use in May 2022 would have meant that the R's AFU was 16 because the R's 17th birthday was not until June 2022. If the R changed the month and year to June 2022, that could be consistent with first use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2021, it would never be possible for the R to have first used at age 17 and also to have first used in June 2021. In this latter situation, BLNTMFU and BLNTYFU would be set to bad data.
Rs entered a new MFU or YFU that differed from what they previously reported. The age based on the revised MFU and YFU (updated in MYR1STBL) still mismatched the AFU, but Rs indicated in the final verification check that the new value from MYR1STBL was correct.	No editing was done in this situation. The CAI program automatically updated AGE1STBL to be consistent with the updated values reported for the MFU and YFU.
The consistency check was triggered between AGE1STBL and MYR1STBL. However, Rs answered the first consistency check (BLCC03, regarding whether the value in MYR1STBL was correct) as "don't know" (DK) or "refused" (REF). Rs then exited the consistency check loop without having resolved the inconsistency.	The AFU from AGE1STBL was retained in BLNTAGE, but BLNTMFU and BLNTYFU were set to bad data.

Table B.13 Edits in the Blunts Section Pertaining to Consistency Checks for Initiation (continued)

Issue	Edits Implemented
The consistency check was triggered between AGE1STBL and MYR1STBL. However, Rs entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to BLNTAGE, but BLNTYFU and BLNTMFU were set to bad data.
Rs entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, they failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. Rs also reported either in the first verification check that MYR1STBL was not correct (i.e., BLCC03 = 6) or reported in the second verification check that the AFU from AGE1STBL was correct (i.e., BLCC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: ▪ The default edit was to set BLNTMFU and BLNTYFU to bad data. ▪ As an exception to the default edit, if the final verification check was answered as DK or REF, BLNTMFU and BLNTYFU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, they failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. Rs also reported in the second verification check that neither answer was correct for what was originally captured in AGE1STBL and MYR1STBL (i.e., BLCC04 = 3).	The following edits were implemented for BLNTAGE, BLNTMFU, and BLNTYFU: ▪ The default edit was to set BLNTAGE, BLNTMFU, and BLNTYFU to bad data. ▪ As an exception to the default edit, if the final verification check was answered as DK or REF, BLNTAGE, BLNTMFU, and BLNTYFU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs triggered the initial consistency check between the AFU, MFU, and YFU and reported that the age from MYR1STBL was correct (i.e., BLCC03 = 4). However, they answered the second consistency check (BLCC04) as DK or REF. Consequently, Rs did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for DK or REF were assigned to BLNTAGE, BLNTMFU, and BLNTYFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when the R first used blunts, or whether the MFU and YFU indicated the R's correct age at initiation. Therefore, BLNTAGE, BLNTMFU, and BLNTYFU all were set to missing values.

Table B.14 Edits in the Blunts Section Other Than Those Pertaining to Consistency Checks for Initiation

Issue	Edits Implemented
Respondents (Rs) reported lifetime use of blunts in question BL01, but they did not know or refused to report in BL02 when they last used them.	The blunt recency variable BLNTREC (corresponding to question BL02) was set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). The 30-day frequency BLNT30DY was left as blank.
Rs reported using blunts in the past 30 days (BL02 = 1), but they previously reported never using marijuana (MJEVER = 2). This discrepancy triggered question BL03 (edited variable BLNT30C1) as a consistency check.	▪ If BLNT30C1 = 1 (i.e., Rs have verified that they never used marijuana): – BLNTEVER was set to 4 (No LOGICALLY ASSIGNED). – BLNTREC was set to 81 (NEVER USED BLUNTS Logically assigned). – BLNT30DY was set to 91 (if blank) or 81 (if a value needed to be overwritten). – Initiation variables (BLNTAGE, BLNTYFU, BLNTMFU) were assigned codes of 981, 9981, and 81, respectively. – Other assignment of legitimate skip codes applies, as described in Section 2.4.2. ▪ If BLNT30C1 = 2 (i.e., Rs have verified past month use of blunts), no editing was done to BLNTEVER or BLNTREC. Also, these data in blunts were not used to edit the MJEVER variable for lifetime marijuana use (Section 2.4.1).
Rs reported using blunts, but they previously reported never using marijuana. The check corresponding to BLNT30C1 was not triggered because Rs reported use of blunts that was less recent than the past 30 days.	▪ BLNTEVER was set to a value of 11 (i.e., “bump” by 10) to signify that Rs reported lifetime use of cigars with marijuana in them but reported never using marijuana in the marijuana section, and Rs were not asked to verify which answer was correct. ▪ No editing was done to the recency values in BLNTREC.

Table B.14 Edits in the Blunts Section Other Than Those Pertaining to Consistency Checks for Initiation (continued)

Issue	Edits Implemented
Rs reported using blunts in the past 30 days, but they reported last using marijuana more than 30 days ago but within the past 12 months or more than 12 months ago. The consistency check BL04 (edited variable BLNT30C2) was triggered for these responses.	▪ If BLNT30C2 = 1 (i.e., Rs verified that the marijuana recency is correct): – BLNTREC was set to 19 (More than 30 days ago LOGICALLY ASSIGNED). – BLNT30DY was set to 93 if blank and to 83 (DID NOT USE BLUNTS IN THE PAST 30 DAYS Logically assigned) if the 30-day frequency data needed to be overwritten. – Note, however, that no editing was done to BLNTREC when BLNT30C2 = 1, but the marijuana recency MJREC had been set to a value of 8 or 9 (Used at some point in the past 12 months or Used at some point in the lifetime, respectively), based on edits in the marijuana section (Section 6.2.2.3). The rationale for doing no editing to BLNTREC was that values of 8 or 9 in MJREC indicated potential use of marijuana in the past month. ▪ If BLNT30C2 = 2 (i.e., Rs verified using blunts in the past 30 days), no editing was done to BLNTEVER or BLNTREC. Also, these data in blunts were not used to edit the marijuana recency MJREC (Section 2.4.1).
Rs reported last using blunts more than 30 days ago but within the past 12 months (BL02 = 2), and they reported last using marijuana more than 12 months ago (MJREC = 3).	BLNT30C2 (BL04) was not triggered in this situation because Rs did not report use of blunts in the past 30 days. ▪ BLNTREC was set to a value of 12 (i.e., “bump” by 10), to signify that Rs reported last using blunts more than 30 days ago but within the past 12 months but reported last using marijuana more than 12 months ago, and they were not asked to verify which answer was correct. ▪ The editing procedures continued to set the 30-day frequency BLNT30DY to 93 because Rs were not past month users of blunts.
Rs reported last using blunts more than 30 days ago but within the past 12 months (BL02 = 2), but the marijuana recency MJREC was assigned codes of 8 or 9 (Used at some point in the past 12 months or lifetime, respectively).	No editing was done to the blunts variables when MJREC had values of 8 or 9 because last use of blunts more than 30 days ago but within the past 12 months was not necessarily inconsistent with when Rs last used marijuana.
Question BL07 (edited variable BLNTNOMJ) was skipped because Rs reported in the tobacco section that they never smoked a cigar and reported in the blunts section that they never smoked a blunt.	BLNTNOMJ was set to a value of 91 (NEVER USED CIGARS/BLUNTS). This edit included situations in which BLNTEVER had been set to 4 (No LOGICALLY ASSIGNED) and BLNTREC to 81 (NEVER USED BLUNTS Logically assigned) because Rs confirmed in BLNT30C2 that they never used marijuana.

Table B.14 Edits in the Blunts Section Other Than Those Pertaining to Consistency Checks for Initiation (continued)

Issue	Edits Implemented
Question BL07 was skipped because Rs reported lifetime use of blunts but not in the past 30 days. Rs also reported in the tobacco section that they did not smoke a cigar in the past 30 days (including situations in which Rs reported never smoking a cigar).	BLNTNOMJ was set to a value of 93 (DID NOT USE CIGARS/BLUNTS IN THE PAST 30 DAYS). This edit also was implemented if Rs reported use of blunts less recently than the past 30 days but reported in the tobacco section that they never smoked a cigar. Even though the cigar and blunts data were inconsistent, BLNTNOMJ still was not applicable.
Question BL07 was skipped because Rs reported lifetime use of blunts but not in the past 30 days. Unlike the preceding situation, Rs reported in the tobacco section that they smoked cigars in the past 30 days.	BLNTNOMJ was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). For instance, if Rs reported last smoking cigars with marijuana in them more than 30 days ago, it could logically be inferred that all past month cigar use had to involve use of cigars that did not have marijuana in them.
Question BL07 was skipped, even though Rs reported use of blunts in the past 30 days (BLNTREC = 1). However, Rs reported lifetime use of cigars in the tobacco section but reported not smoking cigars in the past 30 days.	BLNTNOMJ was set to a value of 14, where 14 = USED BLUNTS PAST 30 DAYS/DIDN'T USE CIGARS PAST 30 DAYS.
Question BL07 was skipped, even though Rs reported use of blunts in the past 30 days (BLNTREC = 1). However, Rs reported in the tobacco section that they had never smoked part or all of a cigar.	BLNTNOMJ was set to a value of 24, where 24 = USED BLUNTS PAST 30 DAYS/NEVER USED CIGARS.
Question BL07 was skipped because the cigar recency variable CIGARREC or the blunts recency variable BLNTREC had missing values or "indefinite" values (i.e., used at some point in the past 12 months, or used at some point in the lifetime).	BLNTNOMJ was left as 98 (blank).
The age at first use (AFU) for blunts (edited variable BLNTAGE) was inconsistent with the R's current age despite the R being prompted to change the AFU.	The final age was accepted as the standard, and BLNTAGE was set to bad data.
BLNTAGE had a value of "don't know" (DK) or "refused" (REF), including situations where this assignment has been made from the consistency check data.	If the questions pertaining to BLNTYFU and BLNTMFU were skipped because Rs answered the AFU as DK or REF, the DK or REF value from BLNTAGE was propagated onto BLNTYFU and BLNTMFU. This edit was designed to indicate the reason that BLNTYFU and BLNTMFU had been skipped. In addition, because Rs may have first used blunts within 1 year of their current age, the month- and year-of-first-use (MFU and YFU) questions may have been relevant to Rs. If Rs had answered the MFU and YFU questions but BLNTAGE had a final code for DK or REF (i.e., due to a consistency check response), data in BLNTYFU and BLNTMFU were overwritten with the corresponding DK or REF value from BLNTAGE. Retaining the MFU and YFU data in this situation would imply that Rs first smoked blunts within 1 year of their current age.

Table B.14 Edits in the Blunts Section Other Than Those Pertaining to Consistency Checks for Initiation (continued)

Issue	Edits Implemented
The MFU was skipped because Rs answered the YFU as DK or REF.	The DK or REF value from BLNTYFU was propagated onto BLNTMFU. That is, if Rs did not know in what year they first used blunts, it was assumed that they would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year.
Rs had missing data for the MFU. However, the AFU and YFU for blunts logically restricted the MFU to a single possible month.	The MFU was logically inferred to be the only month when Rs could have initiated use of blunts.

Table B.15 Edits in the Prior Substance Use Section Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables

Issue	Edits Implemented
Respondents (Rs) reported an age of last use (ALU) that was lower than the age at first use (AFU) from the corresponding earlier substance use section (e.g., MJAGE for marijuana). If the ALU for a given substance, or daily cigarette use, was lower than the corresponding AFU, a consistency check was triggered. However, the consistency check did not allow Rs to change the AFU from the earlier substance use section. The only way they could resolve the inconsistency was by making the ALU consistent with the AFU. Thus, Rs could indicate that the ALU was correct as reported, or they could enter a new ALU that still was inconsistent with the AFU.	The ALU was set to bad data, even if Rs indicated in the consistency check that the previously reported ALU was correct. Thus, as a partial exception to the principle of not editing across sections (Section 2.4.1), AFU data could be used to edit the ALU in the prior substance use section, but not vice versa.
Rs indicated in the final verification check (e.g., LUCG08 for cigarettes) that ALU based on the new month and year of last use (MLU and YLU) was correct (e.g., LUCG08 = 4). The computer-assisted interviewing (CAI) program updated the ALU (such as AGELSTCG for cigarettes) with the value of the age calculated from the MLU and YLU (referred to subsequently as the MYRLST age, such as MYRLSTCG, for cigarettes). However, the new value for the ALU indicated that Rs were more than 1 year younger than their current age at the time they last used the substance (e.g., the R was 16, reported last use of the substance at age 15, but then confirmed an MLU and YLU that meant the R was 14 when the R last used the substance). Had the R initially reported this ALU, the R would not have been routed to the MLU and YLU questions.	The updated value was retained for the ALU (e.g., last use at age 14 for a 16-year-old R in this example). Based on this updated ALU, it was logically inferred that the R should have skipped the MLU and YLU items. A code of 9989 was assigned to the YLU variable (e.g., CIGYLU for cigarettes), and a code of 89 was assigned to the MLU variable (e.g., CIGMLU) (Section 2.4.2).
The final verification check (e.g., LUCG08 for cigarettes) was skipped because Rs entered revised data for the MLU and YLU that made them consistent with the ALU.	No editing was done because Rs were considered to have resolved the inconsistency.
The final verification check (e.g., LUCG08) was skipped because Rs entered a new MLU that was the same as their birth month.	The new MLU could be consistent with the ALU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the ALU, MLU, and YLU, as long as the revised MLU and YLU were potentially consistent with the ALU. However, the MLU and YLU were set to bad data if they could never be consistent with the ALU. Suppose, for example, that a hypothetical R was born in June 2005, was interviewed in March 2023 (age 17 at the time of the interview), reported last use of cigarettes at age 17, and initially reported last use in May 2022. Last use in May 2022 would have meant that the R's ALU was 16 because the R's 17th birthday was not until June 2022. If the R changed the month and year to June 2022, that could be consistent with last use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2021, it would never be possible for the R to have last used at age 17 and also to have last used in June 2021. In this latter situation, CIGMLU and CIGYLU would be set to bad data.

Table B.15 Edits in the Prior Substance Use Section Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables (continued)

Issue	Edits Implemented
The consistency check was triggered between the ALU and the MYRLST age. However, Rs answered the first consistency check (e.g., LUCG05 for cigarettes, regarding whether the value in the MYRLST age was correct) as “don’t know” (DK) or “refused” (REF). Rs then exited the consistency check loop without having resolved the inconsistency.	The value for the ALU was retained, but the MLU and YLU were set to bad data.
The final verification check (e.g., LUCG08 for cigarettes) was skipped because Rs entered the same values for the MLU and YLU that triggered the inconsistency with the ALU in the first place.	The value for the ALU was retained, but the MLU and YLU were set to bad data.
Rs entered a new MLU or YLU that differed from what they previously reported. The MYRLST age based on the revised MLU and YLU still mismatched the ALU, but Rs indicated in the final verification check that this MYRLST age was correct.	No editing was done in this situation. The CAI program automatically updated the ALU to be consistent with the updated values reported for the MLU and YLU.
The consistency check was triggered between the ALU and the MYRLST age calculated from the MLU and YLU. However, Rs answered the first consistency check (e.g., LUCG05 for cigarettes) as DK or REF. Rs then exited the consistency check loop without having resolved the inconsistency.	The ALU was retained, but the MLU and YLU were set to bad data.
Rs entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, they failed to resolve the inconsistency between the ALU and the MYRLST age. Rs also reported either in the first verification check that the MYRLST age was not correct (e.g., LUCG05 = 6 for cigarettes) or reported in the second verification check that the ALU was correct (e.g., LUCG06 = 2 for cigarettes).	No editing was done to the ALU. The following edits were implemented for the MLU and YLU: The default edit was to set the MLU and the YLU to bad data. As an exception to the default edit, if the final verification check was answered as DK or REF, the MLU and YLU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, they failed to resolve the inconsistency between the ALU and the MYRLST age. Rs also reported in the second verification check that neither the ALU nor original MYRLST age was correct (e.g., LUCG06 = 3 for cigarettes).	The following edits were implemented for the ALU, MLU, and YLU: The default edit was to set the ALU, MLU, and YLU to bad data. As an exception to the default edit, if the final verification check was answered as DK or REF, the ALU, MLU, and YLU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs triggered the initial consistency check between the ALU, MLU, and YLU and reported that the MYRLST age calculated from the MLU and YLU was correct (e.g., LUCG05 = 4). However, they answered the second consistency check (e.g., LUCG06 for cigarettes) as DK or REF. Consequently, Rs did not have an opportunity to correct the inconsistency between the ALU, MLU, and YLU.	The relevant codes for DK or REF were assigned to the ALU, MLU, and YLU. The rationale for this edit is that conclusive information did not exist regarding whether the ALU indicated the R’s correct age when the R last used a substance or whether the MLU and YLU indicated the R’s correct ALU. Therefore, the ALU, MLU, and YLU all were set to missing values.

Table B.16 Edits in the Prior Substance Use Section Pertaining to the Age, Year, and Month of Last Use (Other Than for Consistency Checks)

Issue	Edits Implemented
The age of last use (ALU) for a given substance or behavior (MRJAGLST for marijuana, CIGDLLST for daily cigarette smoking) was greater than the respondent's (R's) current age.	The ALU was set to bad data.
The age at first use (AFU) from the corresponding earlier substance use section (e.g., MJAGE for marijuana) had a missing value, but the ALU (e.g., MRJAGLST) was defined. That included situations in which the AFU was set to bad data as part of the "flag and impute" edits (Section 6.2.2.4).	No editing was done to the ALU, as long as it was consistent with the R's age (see above). In addition, no editing was done to the ALU if the imputed AFU (e.g., IRMJAGE for marijuana) was given a value greater than the ALU. (Imputation of the AFU variables did not take into account data from the corresponding ALU variables.)
The edited recency from the earlier substance use section (e.g., CIGREC for cigarettes) indicated that the last use was more than 12 months ago, but the ALU indicated last use at the R's current age. The answer to the ALU would suggest use in the past 12 months. (For cigarettes, either CIGAGLST or CIGDLLST could be inconsistent with the recency.)	No editing was done to the ALU. However, a standard codebook footnote was included for these variables to alert analysts to the fact that values in this section could be inconsistent with values from earlier sections of the interview.
The recency from the earlier substance use section was set to an "indefinite" value of 8 or 9 (used at some point in the past 12 months or used at some point in the lifetime, respectively). These indefinite recency values could be imputed to past month use. If the unedited recency in the corresponding substance use section indicated use more than 30 days ago, Rs would be routed to the prior substance use section and could provide data for the ALU. The latter would suggest that Rs were not past month users, even though the final imputed recency might indicate that they were past month users.	No editing was done to the ALU. The standard codebook footnote described above also applied to this issue.
For Rs who last used substances more than 30 days ago but within the past 12 months, the number of days that they could have used a substance (e.g., marijuana) in the past 12 months (e.g., based on MRJAGLST for marijuana) was less than the number of days in the edited or imputed 12-month frequency for marijuana (e.g., MJYRTOT or IRMJFY, respectively).	In this example, no editing was done to MRJAGLST. The standard codebook footnote described above also applied to this issue.
The ALU variable had a missing value, but the corresponding AFU indicated that Rs first used (or started smoking cigarettes daily) at their current age.	- The ALU was set to equal the AFU. Logically, last use cannot be any later than the R's current age. - As part of this same edit, if the year of first use (YFU) indicated that Rs first used in the year of the interview, then Rs were logically inferred to have last used in the current year.
The year of last use (YLU) had a missing value. In addition, Rs reported first using a substance (or first smoking cigarettes daily) at an age that was 1 year younger than their current age. Rs also reported first using the substance (or first smoking cigarettes daily) in the current year.	No editing was done to the ALU, but the YLU was set to equal the YFU. Logically, if the first use was in the current year, then the last use had to be in the current year as well.

Table B.16 Edits in the Prior Substance Use Section Pertaining to the Age, Year, and Month of Last Use (Other Than for Consistency Checks) (continued)

Issue	Edits Implemented
The month of last use (MLU) had a missing value, but the corresponding month of first use (MFU, or first daily cigarette use) was in the calendar month prior to the interview month.	The MLU was set to equal the MFU. Logically, if Rs were routed to the prior substance use section because they had not used a given substance in the past 30 days, then they also had to have last used that substance in the month prior to the interview month.
The YLU had a missing value or a value that was inconsistent with the YFU. However, the YFU and interview date logically constrained the YLU and MLU to have the same values as the YFU and MFU, respectively.	The YLU and MLU were set to equal the YFU and MFU, respectively.
The ALU had a missing value, but the calculated ALU variable (MYR; e.g., MYRLSTMJ for marijuana) from the YLU, MLU, and birth date was consistent with the AFU and the R's current age. In addition, none of the above edits had been applied.	The ALU was set equal to the calculated ALU (MYR).
Rs gave a valid value for the YLU and reported last use in the same month when they had a birthday, or else they answered the MLU question as "don't know" or "refused."	Two possible values for the ALU were calculated, based on the YLU that Rs reported and their birth year. If neither of these two possible values matched the ALU that Rs reported, then the YLU and MLU were set to bad data.
Rs reported an ALU (and also may have reported values for the YLU and MLU) for when they last used any hallucinogen. However, the ALU, YLU, and MLU variables were skipped for LSD, PCP, or Ecstasy or Molly because Rs had used only LSD, PCP, or Ecstasy or Molly, respectively.	Values from the ALU, YLU, and MLU for any hallucinogen (HALLAGLST, HALLYRLST, and HALLMOLST, respectively) were transferred to the edited variables that had been skipped for LSD, PCP, or Ecstasy or Molly. For example, if Rs were lifetime users only of Ecstasy or Molly, data from HALLAGLST, HALLYRLST, and HALLMOLST were transferred to the corresponding variables ECSAGLST, ECSYLU, and ECSMLU for Ecstasy or Molly.
Rs reported last using any cocaine and crack cocaine in the same year. However, Rs also reported last using any cocaine in a month that was earlier than the month when they reported first using crack cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use).	The MLU for any cocaine was set to bad data. Similar edits were implemented for the MLU variables for any cigarette use as applicable.
The ALU for crack cocaine was later than the corresponding ALU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use).	The ALU for cocaine was set to equal the ALU for crack cocaine. Similar edits were implemented for the ALU any cigarette use as applicable.
The ALU for any cocaine had a missing value. However, Rs reported last using crack cocaine at their current age. This issue also could occur for daily cigarette use (relative to any cigarette use).	- The ALU for any cocaine was set to equal the ALU for crack cocaine. Logically, if Rs last used crack at their current age, then that also had to be the age at which they last used any cocaine. - As part of this same edit, if the crack YLU variable indicated that Rs last used crack in the current year, then the cocaine YLU was set to the current year. Similar edits were implemented for the ALU for any cigarette use, as applicable.

Table B.16 Edits in the Prior Substance Use Section Pertaining to the Age, Year, and Month of Last Use (Other Than for Consistency Checks) (continued)

Issue	Edits Implemented
The YLU for crack cocaine was later than the YLU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use).	The YLU for any cocaine was set to equal the YLU for crack cocaine. Values in the MLU for crack cocaine also were carried over to the MLU for any cocaine. Similar edits were implemented for the YLU and MLU variables for any cigarette use, as applicable.
The YLU for cocaine had a missing value. However, Rs indicated that they last used crack cocaine at an age that was 1 year younger than their current age. Rs also reported last using crack cocaine in the current year. This issue also could occur for daily cigarette use (relative to any cigarette use).	No editing was done to the ALU for any cocaine, but the YLU for any cocaine was set to the current year. Logically, if the last use of crack cocaine was in the current year, then the last use of any cocaine had to be in the current year as well. Similar edits were implemented for the YLU and MLU variables for any cigarette use, as applicable.
Rs last used any cocaine and crack cocaine at the same age and in the same year. However, Rs reported last using crack cocaine in a month that was later than what they reported for last use of any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use).	The MLU for any cocaine was set to equal the MLU for crack cocaine. Similar edits were implemented for the MLU variables for any cigarette use, as applicable.
The cocaine MLU had a missing value, but Rs reported last using crack cocaine in the calendar month prior to the interview month. This issue also could occur for daily cigarette use (relative to any cigarette use).	The MLU for any cocaine was set to the calendar month prior to the interview month. Logically, if Rs were asked the questions about any cocaine and crack cocaine in the prior substance use section because they had not used these substances in the past 30 days, then they also had to have last used any cocaine in the month prior to the interview month. Similar edits were implemented for the MLU variables for any cigarette use, as applicable.
The ALU for any cocaine had a missing value, but the calculated value in the MYR variable for crack cocaine was consistent with both the crack ALU and the R's current age. In addition, none of the above edits was applied. This issue also could occur for daily cigarette use (relative to any cigarette use).	The ALU for any cocaine was set to equal the MYR value for crack cocaine. Similar edits were implemented for the ALUs for any cigarette use, as applicable.
Rs reported last using any hallucinogen and LSD, PCP, or Ecstasy or Molly in the same year. However, Rs also reported last using any hallucinogen in a month that was earlier than the month when they reported first using LSD, PCP, or Ecstasy or Molly.	The MLU for any hallucinogen was set to bad data.

Table B.16 Edits in the Prior Substance Use Section Pertaining to the Age, Year, and Month of Last Use (Other Than for Consistency Checks) (continued)

Issue	Edits Implemented
At least one of the ALUs for LSD, PCP, and Ecstasy or Molly was greater than the ALU for any hallucinogen.	▪ The hallucinogen ALU was set to the latest ALU from the ALUs for LSD, PCP, or Ecstasy. ▪ As part of this same edit, if a given ALU for LSD, PCP, or Ecstasy was moved over to the ALU for any hallucinogen, and the corresponding YLU for LSD, PCP, or Ecstasy was later than the YLU for any hallucinogen, then the later YLU was moved over to the YLU for any hallucinogen. If two or all three of these substances had the same latest ALU value relative to the hallucinogen ALU, then the latest YLU from these substances was moved over to the YLU for any hallucinogen. ▪ If the YLU for LSD, PCP, or Ecstasy was moved over to the YLU for any hallucinogen, and the hallucinogen MLU had a nonmissing value that was earlier than the corresponding MLU for LSD, PCP, or Ecstasy, then the MLU for any hallucinogen was set to the latest MLU from LSD, PCP, or Ecstasy.
The ALU for any hallucinogen had a missing value, but one or more ALUs for LSD, PCP, or Ecstasy indicated that Rs last used at their current age.	▪ The hallucinogen ALU was set to be equal to the R's current age. ▪ As part of this same edit, if a given YLU for LSD, PCP, or Ecstasy indicated that Rs last used in the current year, then the hallucinogen YLU was set to the current year.
The ALU for any hallucinogen had a missing value, and one or more ALUs for LSD, PCP, or Ecstasy indicated that Rs last used at a year younger than their current age. At least one YLU for LSD, PCP, or Ecstasy indicated that Rs last used any of these substances in the current year.	The hallucinogen YLU was set to the current year. No editing was done to the hallucinogen ALU.
The MLU for any hallucinogen had a missing value, but Rs reported last using LSD, PCP, or Ecstasy in the calendar month prior to the interview month.	The MLU for any hallucinogen was set to the calendar month prior to the interview month. Logically, if Rs were asked the questions about any hallucinogen and LSD, PCP, or Ecstasy because they had not used these substances in the past 30 days, then they also had to have last used any hallucinogen in the month prior to the interview month.

Table B.17 Edits Pertaining to the Alcohol and Drug Treatment Variables

Issue	Edits Implemented
Questions in the alcohol and drug treatment section had patterned responses, such as respondents (Rs) keying answers of "1" to all questions that they were asked (Section 2.3.2).	All nonblank values in the edited variables were replaced with codes for "bad data." Variables corresponding to questions that were skipped retained codes of "blank."
Rs specified a source of inpatient substance use treatment in the past 12 months in either SUTINOPLSP (inpatient "OTHER, Specify" question) or SUTOUTOPLSP (outpatient "OTHER, Specify" question). However, these Rs did not report receiving treatment in that location in questions TXSBIPTA through TXSBIPTC.	▪ If the inpatient location was specified in SUTINOPLSP, then the edited variable corresponding to the inpatient location was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP)]. ▪ If the inpatient location was specified in SUTOUTOPLSP, then the edited variable corresponding to the inpatient location was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTOUTOPLSP)].
Rs specified a source of outpatient substance use treatment in the past 12 months in either SUTOUTOPLSP (outpatient "OTHER, Specify" question) or SUTINOPLSP (inpatient "OTHER, Specify" question). However, these Rs did not report receiving treatment in that location in questions TXSBOPTA through TXSBOPTF.	▪ If the outpatient location was specified in SUTOUTOPLSP, then the edited variable corresponding to the outpatient location was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from SUTOUTOPLSP)]. ▪ If the outpatient location was specified in SUTINOPLSP, then the edited variable corresponding to the outpatient location was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP)].
Rs specified in either SUTINOPLSP (inpatient "OTHER, Specify" question) or SUTOUTOPLSP (outpatient "OTHER, Specify" question) that they received substance use treatment via telehealth (over the phone or through video) or while they were in a prison, jail, or juvenile detention center. However, these Rs did not report receiving these sources of treatment in the corresponding direct questions.	The edited variable corresponding to the specific source of substance use treatment was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP or SUTOUTOPLSP)].
Rs specified in either SUTINOPLSP (inpatient "OTHER, Specify" question) or SUTOUTOPLSP (outpatient "OTHER, Specify" question) that they received treatment in a methadone, buprenorphine, or opioid medication-assisted treatment (MAT) program. However, these Rs did not report receiving MAT for their use of drugs (i.e., opioids) in question TXDRRX.	The corresponding edited variable SUTRXDRG was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP or SUTOUTOPLSP)].
Rs specified in either SUTINOPLSP (inpatient "OTHER, Specify" question) or SUTOUTOPLSP (outpatient "OTHER, Specify" question) that they received other services for their alcohol or drug use from a support group, in an emergency room (ER) or emergency department (ED), from a peer support specialist or recovery coach, or from detoxification services. However, these Rs did not report receiving these other services in the corresponding direct questions.	The edited variable corresponding to the specific source of other alcohol or drug use services was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from SUTINOPLSP or SUTOUTOPLSP)].

Table B.17 Edits Pertaining to the Alcohol and Drug Treatment Variables (continued)

Issue	Edits Implemented
Rs specified a specific substance for which they received inpatient treatment in the past 12 months in "OTHER, Specify" question TXSBIPSP. However, these Rs did not report receiving inpatient treatment for that substance in TXSBSUIA through TXSBSUIK.	▪ If Rs ever used the specific substance, then the edited variable corresponding to the receipt of inpatient treatment for that substance was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from SUTINODGSP1-SUTINODGSP5)]. ▪ If Rs previously reported that they never used the specific substance, then the edited variable corresponding to the receipt of inpatient treatment for that substance was assigned a code of 13 [Yes LOG ASSGN (from SUTINODGSP1-5 but no lifetime use)].
Rs specified a specific substance for which they received outpatient treatment in the past 12 months in "OTHER, Specify" question TXSBOPSP. However, these Rs did not report receiving outpatient treatment for that substance in TXSBSUOA through TXSBSUOK.	▪ If Rs ever used the specific substance, then the edited variable corresponding to the receipt of treatment for that substance was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from SUTOUTODGSP1-SUTOUTODGSP5)]. ▪ If Rs previously reported that they never used the specific substance, then the edited variable corresponding to the receipt of outpatient treatment for that substance was assigned a code of 13 [Yes LOG ASSGN (from SUTOUTODGSP1-5 but no lifetime use)].
Rs reported receiving inpatient or outpatient treatment, but they did not report treatment for the use of any of the specific substances that they had used in their lifetime, nor did they report treatment for their use of some other drug.	▪ If Rs reported use of only one substance in their lifetime and reported that they had never used other substances, then the corresponding edited variable for the receipt of inpatient treatment or outpatient treatment for that specific substance was assigned a code of 7 (Yes LOGICALLY ASSIGNED). ▪ If Rs reported lifetime use of more than one drug, but they reported that they had never used alcohol, then the corresponding edited variable for inpatient treatment or outpatient treatment for other drug use was assigned a code of 17 (Received treatment for unspecified drug). ▪ If Rs reported lifetime use of both alcohol and drugs, then the corresponding edited variable for inpatient treatment or outpatient treatment for other drug use was assigned a code of 19 (Substance not specified).
Rs reported receiving substance use treatment via telehealth or while they were in a prison, jail, or juvenile detention center. Rs who had previously reported lifetime use of only alcohol or only drugs were skipped out of follow-up questions asking whether the substance use treatment was received for alcohol use only, drug use only, or both alcohol and drug use.	▪ If Rs reported lifetime use of alcohol only, then the corresponding edited variables SUTPHVIADB (treatment via telehealth) or SUTPRISADB (treatment in a prison, jail, or juvenile detention center) were assigned a code of 4 (For alcohol use only LOGICALLY ASSIGNED). ▪ If Rs reported lifetime use of drugs only, then the corresponding edited variables SUTPHVIADB or SUTPRISADB were assigned a code of 5 (For drug use only LOGICALLY ASSIGNED).

Table B.17 Edits Pertaining to the Alcohol and Drug Treatment Variables (continued)

Issue	Edits Implemented
Rs reported receiving other services for their alcohol or drug use from a support group, in an ER or ED, from a peer support specialist or recovery coach, or from detoxification services. Rs who had previously reported lifetime use of only alcohol or only drugs were skipped out of follow-up questions asking whether the other services were received for alcohol use only, drug use only, or both alcohol and drug use.	▪ If Rs reported lifetime use of alcohol only, then the corresponding edited variables SUTSUPPADB (for support groups), SUTPEERADB (for peer support specialists or recovery coaches), SUTEREDADB (for emergency rooms or departments), or SUTDETXADB (for detoxification) were assigned a code of 4 (For alcohol use only LOGICALLY ASSIGNED). ▪ If Rs reported lifetime use of drugs only, then the corresponding edited variables SUTSUPPADB, SUTPEERADB, SUTEREDADB, or SUTDETXADB were assigned a code of 5 (For drug use only LOGICALLY ASSIGNED).
Rs did not report that they received substance use treatment in the past 12 months (i.e., treatment as an inpatient or outpatient; via telehealth; in a prison, jail, or juvenile detention center; or from a dedicated MAT facility for their alcohol or drug use). Rs also did not report that they received other services for their substance use in the past 12 months (i.e., from a support group, in an ER or ED, from a peer support specialist or recovery coach, or from detoxification services). These Rs were asked whether they sought substance use treatment (TXSBSEK) or whether they thought they should get treatment (TXSBGET) if they did not report that they sought treatment. However, Rs answered “don’t know” (DK) or “refused” (REF) to some questions about substance use treatment. Consequently, it could not be determined unambiguously that these Rs did not receive substance use treatment in the past 12 months.	Corresponding edited variables SUTSEEKTX and SUTSHLDTX had responses of 1 (Yes) and 2 (No), “bumped” by a value of 10. Final values for the variables were 11 [Yes (unknown receipt of treatment)] and 12 [No (unknown receipt of treatment)]. This edit allowed analysts to decide whether to use these data in an analysis.
Rs did not report that they received substance use treatment in the past 12 months (i.e., treatment as an inpatient or outpatient; via telehealth; in a prison, jail, or juvenile detention center; or with MAT for their alcohol or drug use). Rs did report that they received other services for their substance use in the past 12 months (i.e., from a support group, in an ER or ED, from a peer support specialist or recovery coach, or from detoxification services). These Rs were asked whether they sought additional treatment for their substance use (TXSBSEK2) or whether they thought they should get additional treatment (TXSBGET2) if they did not report that they sought additional treatment. However, Rs answered DK or REF to some questions about substance use treatment. Consequently, it could not be determined unambiguously that these Rs did not receive substance use treatment in the past 12 months.	Corresponding edited variables SUTSEEKANL and SUTSHLDANL had responses of 1 (Yes) and 2 (No), “bumped” by a value of 10. Final values for the variables were 11 [Yes (unknown receipt of treatment)] and 12 [No (unknown receipt of treatment)]. This edit allowed analysts to decide whether to use these data in an analysis.

Table B.18 **Prioritization of Specific Skip Logic Edits in the Consumption of Alcohol Section**

Response Pattern	Edit
Data from the alcohol section indicated that the respondents (Rs) never used alcohol.	A code of 91 (or 991, 9991) was assigned to the edited variables in this section. This edit took precedence over situations in which questions may not have been applicable for other reasons (e.g., if Rs were aged 21 or older).
Even if Rs had used alcohol, data from remaining substance use sections indicated that they had never used marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine.	A code of 91 was assigned to the edited variables corresponding to questions CA08 and CA09, pertaining to alcohol use in combination with other drugs in the past 30 days (edited variables CADRKDRUG through CADRKMETH). The code of 91 had the following meaning in these variables: NEVER USED ALCOHOL OR NEVER USED OTHER DRUGS. Similarly, if Rs had not used illicit drugs and they had been logically inferred to be nonusers of at least one drug, then these edited variables were assigned a code of 81, where 81 = NEVER USED ALCOHOL OR NEVER USED DRUGS Logically assigned.
Data from the alcohol section indicated that Rs had used alcohol but not in the past 30 days.	A code of 93 (or 993, 9993) was assigned to the edited variables in this section to indicate that Rs did not use alcohol in the past 30 days.
Even if Rs had used alcohol in the past 30 days, data from remaining substance use sections indicated that they had used marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine in the lifetime but not in the past 30 days.	A code of 93 was assigned to the edited variables corresponding to questions CA08 and CA09. The code of 93 had the following meaning in these variables: DID NOT USE ALCOHOL OR DID NOT USE DRUGS IN THE PAST 30 DAYS.
Rs had used alcohol in the past 30 days (or was potentially a past month user, based on ALCREC = 8 from the alcohol section), but they were aged 21 or older.	The edited variables corresponding to questions CA02 through CA07SP about underage alcohol use were assigned a code of 99 (LEGITIMATE SKIP) if they were blank or otherwise were assigned a code of 89 (LEGITIMATE SKIP Logically assigned).

Table B.18 **Prioritization of Specific Skip Logic Edits in the Consumption of Alcohol Section (continued)**

Response Pattern	Edit
Rs had used alcohol in the past 30 days (or were potentially past month users) and were aged 12 to 20.	▪ If Rs did not indicate in question CA02B that they drank the alcohol in some other place the last time, then the edited variable CADROTHSP (corresponding to question CA02SP) was assigned a legitimate skip code (i.e., 99 if blank or 89 otherwise). ▪ If Rs indicated in question CA03 (edited variable CABUYFRE) that they paid for the last alcohol that they drank, then edited variables corresponding to questions CA07 (edited variable CAFREWHO) and CA07SP (edited variable CAFRESP) were assigned a legitimate skip code. ▪ If Rs indicated in question CA03 that they paid for the last alcoholic beverage that they drank but they gave money to someone else who bought it (question CA04 = 2, corresponding to the edited variable CAGVMONY), then the edited variables corresponding to questions CA05 (CABUYWHO), CA05A (CABPLACE), and CA05B (CABUNDAG) were assigned a legitimate skip code. ▪ If Rs indicated in question CA03 that they paid for the last alcoholic beverage that they drank and also bought it (CAGVMONY = 1), then the edited variable corresponding to question CA06 (CAGVWHO) was assigned a legitimate skip code. In addition, if Rs reported in CA05 that they bought the alcohol in a store, in a restaurant, in a club, or at an event (CA05 = 1), then CABUNDAGE was assigned a legitimate skip code. If Rs reported in question CA05 that they bought the alcohol from another person (CA05 = 2), then CABPLACE was assigned a legitimate skip code. ▪ If Rs indicated in question CA03 that they did not pay for the last alcohol that they drank, then CAGVMONY, CABUYWHO, CABPLACE, CABUNDAG, and CAGVWHO were assigned a legitimate skip code. ▪ If Rs indicated in question CA03 that they did not pay for the last alcohol that they drank and did not report in question CA07 that they got the last alcoholic beverage some other way, then the edited variable CAFRESP (corresponding to CA07SP) was assigned a legitimate skip code.

Table B.18 Prioritization of Specific Skip Logic Edits in the Consumption of Alcohol Section (continued)

Response Pattern	Edit
Rs had used alcohol in the past 30 days and also had used marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine in that period.	- If Rs reported in question CA08 (edited variable CADRKDRUG) that they did not use other drugs at the same time or within a couple of hours of the last time when they drank alcohol in the past 30 days, then the edited variables corresponding to the drugs in question CA09 (i.e., edited variables CADRKMARJ through CADRKMETH) were assigned legitimate skip codes. - Otherwise (e.g., if CA08 was answered as “yes”), if the corresponding data from earlier illicit drug sections indicated that Rs had not used a particular drug in the past 30 days but they had used some other drug in the past 30 days, then the variable corresponding to the drug that they had not used was assigned a legitimate skip code. Suppose, for example, that Rs used other drugs at the same time or within a couple of hours of when they last used alcohol and they had used marijuana but not cocaine in the past 30 days. In this situation, CADRKCOCN (i.e., use of cocaine in combination with alcohol) was assigned a legitimate skip code.
Rs reported using marijuana, cocaine, heroin, hallucinogens, inhalants, or methamphetamine at the same time or within a couple of hours of when they last used alcohol. However, question CA09 about use of alcohol in combination with specific drugs was skipped because Rs reported use of only one of these illicit drugs in the past 30 days.	The edited variable corresponding to the one illicit drug that Rs had used in the past 30 days was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from skip pattern). For all other drugs that Rs had definitely not used in the past 30 days, the edited variables were assigned legitimate skip codes. For example, if Rs answered question CA08 as “yes,” and the only drug that they reported using in the past 30 days was marijuana, then it could be logically inferred that they used marijuana at the same time or within a couple of hours of when they last used alcohol, and the edited variable CADRKMARJ was assigned a code of 5.
The lifetime binge alcohol question CA10 (for consumption of four or more drinks on a single occasion for females or consumption of five or more drinks on a single occasion for males) had been skipped because either of the following occurred: - The alcohol use data indicated binge alcohol use on 1 or more days in the past 30 days based on the thresholds described above for males and females. - Question CA01 (edited variable CADRLAST) indicated that Rs binge drank the last time that they drank in the past 30 days.	The edited variable CABINGEVR (corresponding to question CA10) was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).
Question CA10 was answered as “no,” indicating that Rs had never engaged in binge drinking (see above).	The edited variables CABINGAGE (corresponding to question CA11), CABINGYFU (derived from questions CA11a through CA11c), and CABINGMFU (derived from questions CA11c and CA11d) were assigned legitimate skip codes.

Table B.19 Edits in the Consumption of Alcohol Section Pertaining to the Binge Alcohol Use Variables

Issue	Edits Implemented
Respondents (Rs) indicated in the final verification check for the age when they first engaged in binge alcohol use (BACC06) that the age at first binge alcohol use based on the new month and year of first use (MFU and YFU) was correct (i.e., BACC06 = 4). The computer-assisted interviewing (CAI) program updated the age at first use (AFU; AGE1STBA) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBA). However, the new value for the AFU indicated that Rs were more than 1 year younger than their current age at the time they first engaged in binge alcohol use. Had Rs initially reported this AFU, they would not have been routed to the relevant MFU and YFU questions.	The updated value was retained in the edited AFU variable CABINGAGE. Based on the updated value for the AFU, it was logically inferred that Rs should have skipped the MFU and YFU items. A code of 9989 was assigned to the relevant YFU variable (CABINGYFU), and a code of 89 was assigned to the corresponding MFU variable (CABINGMFU) (Section 2.4.2).
The final verification check (BACC06) was skipped because Rs entered revised data for the relevant MFU and YFU that made MFU and YFU for first binge alcohol use consistent with the AFU.	No editing was done because Rs were considered to have resolved the inconsistency.
The final verification check (BACC06) was skipped because Rs entered a new MFU that was the same as their birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the revised MFU and YFU were set to bad data if they could never be consistent with the AFU.
Rs entered a new MFU or YFU that differed from what they previously reported. The age based on the revised MFU and YFU (updated in MYR1STBA) still mismatched the AFU, but Rs indicated in the final verification check that the new value from the relevant MYR variable was correct.	No editing was done in this situation. The CAI program automatically updated the value for the AFU to be consistent with the updated values reported for the MFU and YFU.
The consistency check was triggered between the AGE1STBA variable stored by the CAI program and the corresponding MYR1STBA variable. However, Rs answered the first consistency check (i.e., BACC03, regarding whether the value in MYR1STBA was correct) as "don't know" (DK) or "refused" (REF). Rs then exited the consistency check loop without having resolved the inconsistency.	The AFU value that had been stored in the AGE1STBA variable was retained, but the corresponding MFU and YFU variables were set to bad data.
The consistency check was triggered between the AGE1STBA and corresponding MYR1STBA variable. However, Rs entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to the AFU, but the YFU and MFU variables were set to bad data.

Table B.19 Edits in the Consumption of Alcohol Section Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
Rs entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR1STBA variable. However, they failed to resolve the inconsistency between the AGE1STBA variable and the updated value in the MYR1STBA variable. Rs also reported either in the first verification check that MYR1STBA was not correct (i.e., BACC03 = 6) or reported in the second verification check that the AFU from the AGE1STBA variable was correct (i.e., BACC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: - The default edit was to set the MFU and YFU to bad data. - As an exception to the default edit, if the final verification check was answered as DK or REF, the MFU and YFU were assigned the DK or REF code that corresponded to the answer in the final verification check.
Rs entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR1STBA variable. However, they failed to resolve the inconsistency between the AGE1STBA variable and the updated value in the MYR1STBA variable. Rs also reported in the second verification check that neither answer was correct for what was originally captured in the AGE1STBA and MYR1STBA variables (i.e., BACC04 = 3).	The following edits were implemented for the AFU, MFU, and YFU: - The default edit was to set the AFU, MFU, and YFU to bad data. - As an exception to the default edit, if the final verification check was answered as DK or REF, the AFU, MFU, and YFU were assigned the DK or REF code that corresponded to the answer in the final verification check.
The AFU was inconsistent with the value from the MYR1STBA variable (based on the MFU and YFU). Rs also reported that the age from the MYR1STBA variable was correct (i.e., BACC03 = 4). However, they answered the second consistency check (BACC04) as “don’t know” or “refused.” Consequently, Rs did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for DK or REF were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R’s correct age when the R first engaged in binge alcohol use, or whether the MFU and YFU indicated the R’s correct age at initiation of binge alcohol use. Therefore, the AFU, MFU, and YFU all were set to missing values.
Rs reported first binge alcohol use at an age that was earlier than what was reported for their first use of any alcohol or at an age that was greater than their current age.	The edited AFU variable (i.e., CABINGAGE) was set to bad data. Any nonmissing values in the YFU and MFU also were set to bad data.
Rs reported first binge alcohol use in a year that was earlier than what they reported for the year in which they first used any alcohol.	The edited YFU variable was set to bad data. Any nonmissing values in the MFU also were set to bad data.
Rs reported first binge alcohol use in the same year in which they first used any alcohol. However, Rs reported first binge alcohol use in a month that was earlier than the month in which they first used any alcohol.	The edited MFU variable was set to bad data.
Rs reported first binge alcohol use in the same month as their birth month. However, Rs could never have first engaged in binge alcohol use at the reported age and also have first engaged in binge alcohol use in the year that they reported.	The relevant YFU and MFU variables were set to bad data.

Table B.19 Edits in the Consumption of Alcohol Section Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
The edited AFU variable for binge alcohol use had a missing value. However, Rs first used alcohol at their current age.	▪ The binge alcohol AFU variable with the missing value was set to the R's current age. ▪ The corresponding editing indicator (EI) variable was set to a value of 2, where 2 = Logically assigned data. ▪ If the YFU and MFU had been set to bad data but the unedited values that Rs reported for the YFU and MFU were the same as those in ALCYFU and ALCMFU, then the relevant binge YFU and MFU variables were reset to their unedited values.
The edited YFU variable for binge alcohol use had a missing value. However, Rs first used alcohol in the current survey year.	▪ The binge alcohol YFU variable with the missing value was set to the current year. ▪ If the edited YFU was not the same as the unedited value, then the corresponding EI variable was set to a value of 2, where 2 = Logically assigned data. ▪ If the MFU had a missing value and ALCMFU indicated that Rs first used alcohol in the interview month, then the binge MFU was set to the same value as ALCMFU. If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (i.e., EIBINGMFU) was set to a value of 2.
The edited MFU variable for binge alcohol use had a missing value. However, the YFU data and interview date supplied information on when Rs logically first engaged in binge alcohol use.	▪ The missing value in the MFU was replaced with the definite month in which Rs logically first engaged in binge alcohol use. For example, if the R reported first use of any alcohol and first binge alcohol use in the same year, and the R first used any alcohol in December of that year, then December logically was the only month in that year in which the R could have initiated binge alcohol use. ▪ If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (i.e., EIBINGMFU) was set to a value of 2.

Table B.20 Edits Pertaining to the Special Topics Section

Issue	Edits Implemented
Respondents (Rs) reported never having been arrested or answered the lifetime arrest question as “don’t know” or “refused” but reported being on probation or parole in the past 12 months.	Rs were logically inferred to have been arrested at least once in their lifetime (i.e., BOOKED = 3). The rationale for this edit was that someone could not be on probation or parole without first having been arrested for a crime. The skipped variables pertaining to arrests in the past 12 months retained a value of blank.
Rs reported being arrested in the past 12 months, did not report being arrested for a specific crime in that period, but reported being arrested for that crime as “some other offense.”	Rs were logically inferred to have been arrested for that crime (i.e., assigned a code of 3). No further editing was done to the affirmative answer where Rs reported being arrested for “some other offense” (BKOTH). Similarly, no further editing was done to the “OTHER, Specify” variable (BKOTHOFF) that indicated the crime for which Rs were arrested (Section 2.4.5).
Rs reported being arrested at least once in the past 12 months but initially answered all past year arrest questions as “no.” However, Rs specified an offense in the series of follow-up questions SP03R, SP03RSP, SP03S, and SP03SSP.	The variable for “some other offense” (BKOTH from question SP03Q) was assigned a code of 3 (i.e., Yes LOGICALLY ASSIGNED). The offense that was specified in SP03RSP or SP03SSP was assigned to the edited “OTHER, Specify” variable BKOTHOFF.
Rs reported being arrested at least once in the past 12 months but answered all specific past year arrest questions as “no” and reported nothing in the “OTHER, Specify” variable BKOTHOFF to support the indication of being arrested.	The response was retained to indicate that Rs had been arrested in the past 12 months. A code of 5 (i.e., Offense unknown LOGICALLY ASSIGNED) was assigned to the “some other offense” variable (BKOTH).
Rs reported being arrested for every offense in the past 12 months that was asked about in the section. (For youths aged 12 to 17, that included reports of being arrested for possession of tobacco. This question was skipped for adults.)	The edits differed, depending on what Rs specified for their “other” offense: ▪ If a valid “other” offense was not specified, the entire series of past year offense variables was assigned a bad data code. ▪ If Rs gave a valid response for some other offense for which they were arrested in the past 12 months, the data were retained to indicate that Rs were arrested for this other offense. However, the variables pertaining to arrests for all other offenses were set to bad data. ▪ For adults, the variable pertaining to arrests for possession of tobacco (BKPOSTOB) continued to be assigned a legitimate skip code.
Rs reported being arrested only one time in the past 12 months, did not report being arrested for some other offense (BKOTH = 2), but reported being arrested for every other offense in that same period.	Not including BKOTH or its associated “OTHER, Specify” variable (BKOTHOFF), the variables pertaining to arrests for specific offenses in the past 12 months were assigned a bad data code. For adults, the BKPOSTOB variable continued to be assigned a legitimate skip code.
Rs reported being arrested 80 or more times in the past 12 months.	The variable pertaining to the number of arrests in the past 12 months (NOBOOKYR) was set to bad data.
Rs had alternating “yes/no” or “no/yes” patterns to all questions about arrests for specific offenses in the past 12 months (e.g., SP03a = 1, SP03b = 2, SP03c = 1).	All variables pertaining to arrests for specific offenses in the past 12 months were set to bad data.

Table B.21 Edits Pertaining to the Mental Health Services Utilization Variables

Issue	Edits Implemented
Questions in the mental health services utilization section had patterned responses, such as respondents (Rs) keying answers of "1" to all questions that they were asked (Section 2.3.2).	All nonblank values in the edited variables were replaced with codes for "bad data." Variables corresponding to questions that were skipped retained codes of "blank."
Rs specified a source of inpatient mental health treatment in the past 12 months in either MHTINOPLSP (inpatient "OTHER, Specify" question) or MHTOUTOPLSP (outpatient "OTHER, Specify" question). However, these Rs did not report receiving treatment in that location in questions TXMHIPTA through TXMHIPTC.	- If the inpatient location was specified in MHTINOPLSP, then the edited variable corresponding to the inpatient location was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP)]. - If the inpatient location was specified in MHTOUTOPLSP, then the edited variable corresponding to the inpatient location was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTOUTOPLSP)].
Rs specified a source of outpatient mental health treatment in the past 12 months in either MHTOUTOPLSP (outpatient "OTHER, Specify" question) or MHTINOPLSP (inpatient "OTHER, Specify" question). However, these Rs did not report receiving treatment in that location in questions TXMHOPTA through TXMHOPTF.	- If the outpatient location was specified in MHTOUTOPLSP, then the edited variable corresponding to the outpatient location was assigned a code of 3 [Yes LOGICALLY ASSIGNED (from MHTOUTOPLSP)]. - If the outpatient location was specified in MHTINOPLSP, then the edited variable corresponding to the outpatient location was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP)].
Rs specified that they received mental health treatment by taking prescription medication, via telehealth (over the phone or through video), or while they were in a prison, jail, or juvenile detention center in either MHTINOPLSP (inpatient "OTHER, Specify" question) or MHTOUTOPLSP (outpatient "OTHER, Specify" question). However, these Rs did not report receiving these sources of treatment in the corresponding direct questions.	The edited variable corresponding to the specific source of mental health treatment was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP or MHTOUTOPLSP)]. Because the inpatient and outpatient mental health treatment questions apply to locations where respondents received treatment, it is unlikely for respondents to specify taking prescription medication as "some other place" where they received inpatient or outpatient treatment. If this situation were to occur, the code of 5 would also be used to infer that respondents took medication that was prescribed to them to help with their mental health, emotions, or behavior.
Rs specified that they received other services for their mental health, emotions, or behavior from a support group, in an emergency room (ER) or emergency department (ED), or from a peer support specialist or recovery coach in either MHTINOPLSP (inpatient "OTHER, Specify" question) or MHTOUTOPLSP (outpatient "OTHER, Specify" question). However, these Rs did not report receiving these other services in the corresponding direct questions.	The edited variable corresponding to the specific source of other mental health services was assigned a code of 5 [Yes LOGICALLY ASSIGNED (from MHTINOPLSP or MHTOUTOPLSP)].

Table B.21 Edits Pertaining to the Mental Health Services Utilization Variables (continued)

Issue	Edits Implemented
Rs did not report that they received mental health treatment in the past 12 months (i.e., that they received treatment as an inpatient or outpatient, via telehealth, or in a prison, jail, or juvenile detention center or that they took prescription medication for their mental health). Rs also did not report that they received other services for their mental health, emotions, or behavior in the past 12 months (i.e., from a support group, in an ER or ED, or from a peer support specialist or recovery coach). These Rs were asked whether they sought mental health treatment (TXMHSEK) or whether they thought they should get treatment (TXMHGET) if they did not report that they sought treatment. However, Rs answered “don’t know” (DK) or “refused” (REF) to some questions about mental health treatment. Consequently, it could not be determined unambiguously that these Rs did not receive mental health treatment in the past 12 months.	Corresponding edited variables MHTSEKTX and MHTSHLDTX had responses of 1 (Yes) and 2 (No), “bumped” by a value of 10. Final values for the variables were 11 [Yes (unknown receipt of treatment)] and 12 [No (unknown receipt of treatment)]. This edit allowed analysts to decide whether to use these data in an analysis.
Rs did not report that they received mental health treatment in the past 12 months (i.e., that they received treatment as an inpatient or outpatient, via telehealth, or in a prison, jail, or juvenile detention center or that they took prescription medication for their mental health). Rs did report that they received other services for their mental health, emotions, or behavior in the past 12 months (i.e., from a support group, in an ER or ED, or from a peer support specialist or recovery coach). These Rs were asked whether they sought additional treatment for their mental health (TXMHSEK2) or whether they thought they should get additional treatment (TXMHGET2) if they did not report that they sought additional treatment. However, Rs answered DK or REF to some questions about mental health treatment. Consequently, it could not be determined unambiguously that these Rs did not receive mental health treatment in the past 12 months.	Corresponding edited variables MHTSEKANL and MHRSHLDANL had responses of 1 (Yes) and 2 (No), “bumped” by a value of 10. Final values for the variables were 11 [Yes (unknown receipt of treatment)] and 12 [No (unknown receipt of treatment)]. This edit allowed analysts to decide whether to use these data in an analysis.

Appendix C: Race and Hispanic/Latino Group Codes

C.1 Introduction

For the 2023 National Survey on Drug Use and Health (NSDUH), it was not uncommon for respondents to feel that the race or Hispanic/Latino group given in the questionnaire did not apply to them. In these situations, respondents were given the opportunity to enter a group they felt best described them. These responses were called "OTHER, Specify" responses because they were typed in a part of the question that asked the interviewer or respondent to specify a different response. These "OTHER, Specify" responses were then matched to codes to describe the responses, which were collected and maintained in a file known as a "dictionary." Other-specify responses from each survey year were matched against this file, and any responses without codes were given new codes and added to the dictionary. Consequently, the size of the dictionary file has increased each survey year.

Most of the time, new unmatched responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican." As discussed in Section 4.2.7, many respondents provided a race in the "OTHER, Specify" response to the Hispanic/Latino group question and vice versa, so responses to both questions were examined in the creation of each variable. This appendix details the procedures that were implemented to assign race and Hispanic/Latino values to respondents based on "OTHER, Specify" responses to the questionnaire.

Once a racial group was chosen that represented the "OTHER, Specify" response, this was combined with any group(s) selected in the race and Hispanic/Latino group questions. If the information provided in the "OTHER, Specify" response was so general that formal imputation seemed to be required, and more specific information was available in the selected group(s), then the final assignment of a racial group was made using only the information from the selected group(s) and the "OTHER, Specify" information was ignored.

C.2 Race

Three questions (QD05, QD05PI, and QD05ASIA) focused on the respondent's race. Respondents were permitted to select more than one race in each question. If they selected "Native Hawaiian or Other Pacific Islander" as one of their races in QD05, they were routed to QD05PI, and if they selected "Asian" as one of their races in QD05, they were routed to QD05ASIA. Respondents were only given the option to specify their race for QD05 if they answered "Don't Know." In QD05PI and QD05ASIA, the respondent had the opportunity to select "Other" as the group. They were then prompted to specify their selection by typing the response. The "OTHER, Specify" responses to these three questions were considered together.

The Hispanic/Latino group question (QD04), discussed in Section C.3, also had an "OTHER, Specify" response. Due to the relationship between Hispanic/Latino origin and race reporting, Hispanic/Latino origin was used in the editing of race, and vice versa. In the process of editing race, the "OTHER, Specify" response to the Hispanic/Latino group question (QD04) was consulted (if it existed), provided no race information was identified in QD05, QD05PI, or

QD05ASIA. Similarly, in the process of editing the Hispanic/Latino group, the “OTHER, Specify” responses to the race questions (QD05, QD05PI, and QD05ASIA) were consulted (if they existed), provided no Hispanic/Latino group information was identified in QD04.

C.2.1 Race “OTHER, Specify” Responses

The four types of race “OTHER, Specify” responses can be grouped into four categories: (1) directly mapped codes, (2) indirectly mapped codes, (3) codes informative for imputation, and (4) codes not informative for imputation. Codes that were directly or indirectly mapped were assigned a race group and did not require imputation. Codes that were informative for imputation were used to restrict imputation, and codes that were noninformative resulted in unrestricted imputation. For details on the imputation procedures for race, see Section 4.3.2.

Directly Mapped Codes. Directly mapped codes were codes mapped to one or more of the groups given in the questionnaire. There were two types of directly mapped codes: (1) racial group codes and (2) geographic category codes. Racial group codes were equivalent to one or more of the offered groups in QD05, QD05PI, or QD05ASIA. For example, a response such as “mestizo” mapped directly to two groups in QD05 (White and American Indian/Alaska Native). Geographic category codes corresponded to a country where census data indicated a racially homogeneous society or a country where census data was not available, but a predominant race group(s) could be inferred. For example, an entry of “Mozambique” mapped to “Black/African American” because the Mozambique census data indicated that nearly all Mozambicans were African.

Indirectly Mapped Codes. Codes that were indirectly mapped corresponded to countries where census data were used to map to race groups of racially heterogeneous countries. A racial group from among the 11 groups given in the questionnaire was chosen by generating a random number and allocating the race based on a comparison of the random number with the proportions of races in the geographical entity’s (mostly countries) census. For example, an entry of “St. Thomas” would have a 76 percent chance of being allocated to the Black/African American group, because the census indicated that 76 percent of individuals from the U.S. Virgin Islands were Black. If two or three heterogeneous countries were entered in the “OTHER, Specify” response, the final race was allocated using the following procedure: (1) randomly assign races based on the proportions for each country mentioned, and (2) combine the results. Exceptions to these rules occurred with the groups Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain). Census data were only used for these Hispanic/Latino groups when specified in combination with a country(ies) not appearing in this list of Hispanic/Latino groups. See Section C.2.3.1 for more detail.

Informative Codes for Formal Imputation Procedures. Some “OTHER, Specify” responses could not be definitively mapped to a race group but could be used to limit the final imputation. With these informative codes, the final imputation was restricted according to the information that was available. For example, a response of “mixed” resulted in an imputation among donors with two or more races, and a response of “brown” resulted in an imputation among donors who were not single-race White. That is, logical constraints (see Section 3.3.3.1) were used to restrict the set of donors to ensure imputation-revised values were consistent with

the information provided by the respondent. No imputation was required if more specific information was available from responses QD05, QD05Asia, or QD05PI.

Noninformative Codes. Finally, a noninformative response that was not accompanied by a response to one of the given offered groups resulted in an unrestricted imputation. Religious identifications (e.g., Muslim) were considered noninformative, even if the religion was usually associated with a particular ethnic group (e.g., Shinto is usually associated with Japanese).

[Table C.1](#) lists all the race codes, along with supplementary information related to race codes. Special situations associated with the four types of codes described in this section are discussed in the following sections. For most codes, the final assignment depended upon whether the response was given in QD05, QD05PI, or QD05ASIA. For informative codes described above, the six Hispanic/Latino codes—Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain)—were treated differently depending upon whether they were listed in conjunction with other racial or geographical entities.

Codes with an asterisk were those that caused the Hispanic/Latino indicator to be edited to a “yes.” That is, if QD03 was either missing or “no,” and any of these codes appeared as an “OTHER, Specify” response to QD05, QD05PI, or QD05ASIA, the edited Hispanic/Latino indicator (EDHOIND) was set to 1 to denote “Hispanic/Latino,” and the imputation indicator for the Hispanic/Latino indicator (IIHOIND) was set to 2 to indicate “logically assigned.” See Section 4.2.7.4 for more details on the edited Hispanic/Latino indicator. Note that EDHOIND also could be edited to a “no.” This is discussed in Section C.3.1.

C.2.1.1 Handling of Directly Mapped Codes

For codes that were directly mapped, the final mapped race code is indicated in the “Results” columns of [Table C.1](#). Usually, the handling of directly mapped codes that were racial groups, Other Pacific Islander geographic categories, or Asian geographic categories did not depend upon whether the response was observed in QD05, QD05PI, or QD05ASIA. The exceptions to this rule occurred if the response included a reference to “Indian,” which was mapped to “American Indian/Alaska Native” if the response was given in QD04 or QD05 and to “Asian Indian” if the response was given in QD05ASIA or QD05PI. On the other hand, for directly mapped codes that were non-Other Pacific Islander or non-Asian geographic categories, the final mapping always depended upon which question the response was observed. In this instance, if the code was observed in QD05PI, it was always mapped to “Other Pacific Islander,” and if the code was observed in QD05ASIA, the code was always mapped to “Other Asian.”

C.2.1.2 Handling of Indirectly Mapped Codes

Most of the time indirectly mapped codes refer to heterogeneous countries where census data were used. In these situations, the race was assigned by comparing a randomly generated number to the proportion of each racial group in that country’s census. The final mapping of the indirectly mapped codes also depended upon whether the response was in QD05, QD05PI, or QD05ASIA. If the code was observed in QD05ASIA, it was mapped to “Other Asian,” provided none of the entries observed were Asian racial groups, Asian countries, or countries with a

specific Asian minority. Similarly, if the code was observed in QD05PI, it was mapped to “Other Pacific Islander.” Codes where there was at least one Asian minority in a specified heterogeneous country that was not all Asian, and the response was given in QD05ASIA, were handled on a case-by-case basis. The resulting strategy was either a direct mapping or a different indirect mapping than that given if the response was in QD05.

When census data were used, it was not uncommon to find that a small proportion of the population was identified as “Other.” In the rare instance that the randomly generated number indicated the respondent belonged to this “Other” group, then the selected race was determined by imputation. Codes where this was possible are identified with a superscript “I” in the “Type” columns of [Table C.1](#).

Generally, if two entries (countries or racial groups) were observed, the race for each entity was determined first (either through a direct map or a random assignment using census data), and then the two races were combined. Sometimes a racial group was listed along with a geographical entity. As stated earlier, most of the time the geographical entity was ignored because it was usually assumed that the respondent was a resident of the listed country who also happened to be identified with the given racial group. However, it became clear on occasion that the respondent had parentage that belonged to the racial group and different parentage that came from the listed country. In these instances, the racial group was treated in the same manner as a homogeneous country of that race, and the determination of a final race was conducted in the same manner as if two countries had been listed. If one of the races listed was an Asian racial group, for example, then the response was treated in the same manner whether it was observed in QD05 or QD05ASIA. If the final assignment depended upon the census data of two indirectly mapped codes or an indirectly mapped code and a racial group, “double census” is parenthetically indicated in the third column of [Table C.1](#). If three indirectly mapped codes were indicated by the respondent, “triple census” is indicated.¹

Details about how to handle census information for each indirectly mapped code are shown in [Table C.2](#). Note that the racial groups for each country listed in [Table C.2](#) have been modified to conform to the racial groups specified by the questionnaire. For example, the Black racial group from other countries has been modified to the Black/African American group. Every category and restricted imputation level with a nonzero probability of selection is listed. If a code had an indirect map (using census data) for QD05, but had a direct map for QD05ASIA, this is not specified in [Table C.2](#). Instead, this information must be obtained from [Table C.1](#). Explanations of the categories that are not self-explanatory are listed below.

¹ When an indirectly mapped code with superscript I appeared as a component in a double census or triple census code, the probability associated with the “other” category was distributed among the races appearing in the census. This was the simplest way to preserve race information from all the component indirectly mapped codes. A more complicated alternative would be to impute race information for each component country, even if the “other” category was selected at random for one or more of the race groups in the census. See the entry for “Costa Rica” (code 78) in [Table C.2](#) for an example.

- White or Mestizo: Imputation was restricted to respondents who were either White or Mestizo (i.e., White and American Indian/Alaska Native only). See Section 4.2.7.3.2 for the explanation of level 34 of EDACE.
- Not American Indian: Imputation was restricted to respondents who were a single race other than American Indian/Alaska Native or were multiple race and American Indian/Alaska Native was not one of their component races. See Section 4.2.7.3.2 for the explanation of level 35 of EDACE.
- Multiple: Imputation was restricted to respondents who were multiple race. See Section 4.2.7.3.2 for the explanation of level 32 of EDACE.

C.2.1.3 Handling of Codes Informative for Imputation

For six Hispanic/Latino codes that were highly prevalent in the data, census data were not used to assign the final racial group (i.e., the six groups listed in QD04). Instead, the final racial group for respondents who responded “Mexican,” “Puerto Rican,” “Central or South American,” “Cuban,” “Dominican,” or “Spanish” was determined by a restricted imputation with donors who indicated one of these groups in QD04. Furthermore, if a respondent indicated any combination of these six groups, the final racial group was determined using a restricted imputation with donors who were from the geographical entities listed. However, if any of these six Hispanic/Latino groups was listed along with a second code that was not among the six racial groups, census data were used along with the census data from the second country listed. More details about how specific restricted imputations were conducted are shown in [Table C.3](#).

C.2.1.4 Handling of Noninformative Codes

For noninformative codes, a final race could still have been assigned based on responses to other groups in QD05. If no other groups were listed in QD05, race was imputed, where the imputation was restricted to a Hispanic/Latino group if the respondent gave Hispanic/Latino information in QD04. Otherwise, the final race was determined through an unrestricted imputation.

C.3 Hispanic/Latino Origin

As with the race questions, Hispanic/Latino respondents² had the opportunity to specify a Hispanic/Latino group by giving the response “Other” to QD04, the Hispanic/Latino group question. Also, respondents were permitted to select multiple Hispanic/Latino groups in response to QD04.

The QD05, QD05PI, and QD05ASIA questions are discussed in Section C.2. They also have “OTHER, Specify” responses, which were gleaned for Hispanic/Latino group information whenever no Hispanic/Latino group information was available from QD04.

² For the purposes of the instrument question-routing, Hispanic/Latino respondents were identified by their response to question QD03: “Are you of Hispanic, Latino, or Spanish origin or descent?”

C.3.1 Hispanic/Latino Group “OTHER, Specify” Responses

There were only two types of Hispanic/Latino group “OTHER, Specify” responses: (1) those that mapped to one or more EDQD04xx² variables and (2) those that were ignored. There were no census-based routines and no responses that caused the imputation to be restricted. The imputation of a Hispanic/Latino group was restricted only when race information was available.

[Table C.4](#) lists all the Hispanic/Latino group codes used in the 2023 survey and the Hispanic/Latino groups to which they mapped. Note that these mappings utilized the arbitrary priority rule provided in Section 4.2.7.4.3. This rule was used to create EDHOGRP, which skipped the intermediate step of recording the Hispanic/Latino groups indicated in QD04. These are recorded in the EDQD04xx variables, which are described in Section 4.2.7.4.2, along with the creation of EDHOGRP. The Hispanic/Latino code 600, “Stated Clearly as Not Hispanic/Latino,” was unique in that it could be used to edit the Hispanic/Latino indicator, if needed. If QD03 was missing or 1, then EDHOIND was edited to a 2 if this code appeared in QD04, QD05, QD05PI, or QD05ASIA.

³ See Section 4.2.7.4.2 for a discussion of EDQD04xx.

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
21	White	Direct Map	White	Direct Map	White	Direct Map	White
22	Black/African American	Direct Map	Black/African American	Direct Map	Black/African American	Direct Map	Black/African American
23	American Indian/ Alaska Native	Direct Map	American Indian/ Alaska Native	Direct Map	American Indian/ Alaska Native	Direct Map	American Indian/ Alaska Native
24	Native Hawaiian	Direct Map	Native Hawaiian	Direct Map	Native Hawaiian	Direct Map	Native Hawaiian
25	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander
26	Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian
27	Chinese	Direct Map	Chinese	Direct Map	Chinese	Direct Map	Chinese
28	Filipino	Direct Map	Filipino	Direct Map	Filipino	Direct Map	Filipino
29	Japanese	Direct Map	Japanese	Direct Map	Japanese	Direct Map	Japanese
30	Korean	Direct Map	Korean	Direct Map	Korean	Direct Map	Korean
31	Vietnamese	Direct Map	Vietnamese	Direct Map	Vietnamese	Direct Map	Vietnamese
32	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
33	Asian (nonspecific)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
34	Guamanian/Chamorro	Direct Map	Guamanian or Chamorro	Direct Map	Guamanian or Chamorro	Direct Map	Guamanian or Chamorro
35	Indian (Asian or American unclear)	Direct Map	American Indian/ Alaska Native	Direct Map	Asian Indian	Direct Map	Asian Indian
36	Other Pacific Islander (nonspecific)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
50	Belize	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
51	Guyana	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
52	Suriname	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
53	Haiti	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
54	Trinidad and Tobago	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
55	Jamaica	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
56	Virgin Islands (St. Thomas, St. Croix)	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
57	Bahamas	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
58	Barbados	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
59	Grenada	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
60	St. Lucia	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
61	St. Vincent & the Grenadines	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
62	Dominica	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
63	Other West Indies	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
64	Brazil	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Japanese
65	Canada	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
66	Bahamas & Haiti	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
67	Brazil & Portugal	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
68	Jamaica & West Indies	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
70	Mexico	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
71	Puerto Rico	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
72	Cuba	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
73	Dominican Republic	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
74	Guatemala	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
75	Honduras	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
76	El Salvador	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
77	Nicaragua	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
78	Costa Rica	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Chinese
79	Panama	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
80	Colombia	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
81	Venezuela	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
82	Ecuador	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
83	Peru	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Japanese
84	Bolivia	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
85	Chile	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
86	Argentina	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
87	Paraguay	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
88	Uruguay	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
89	Mexico & Puerto Rico	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
90	Mexico & Cuba	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
91	Mexico & Dominican	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
92	Mexico & Spain	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
93	Puerto Rico & Cuba	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
94	Puerto Rico & Dominican	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
95	Puerto Rico & Spain	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
96	Cuban & Dominican	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
97	Cuban & Spain	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
98	Dominican & Spain	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
100	Norway	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
101	Sweden	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
102	Denmark	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
103	United Kingdom	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
104	Ireland	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
105	Portugal	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
106	Spain	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
107	Germany	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
108	France	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
109	Italy	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
110	Netherlands	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
111	Belgium	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
112	Greece	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
113	Russia	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
114	Ukraine	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
115	Turkey	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
116	Other Western Europe	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
117	Other Eastern Europe	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
118	Other Southern Europe	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
119	Morocco	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
120	Algeria	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
121	Tunisia	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
122	Libya	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
123	Egypt	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
124	Other North Africa	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
125	Saudi Arabia	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
126	Yemen	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
127	Oman	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
128	UAE	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
129	Qatar	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
130	Bahrain	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
131	Israel	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
132	Iraq	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
133	Kuwait	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
134	Iran	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
135	Other Middle East	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
136	Armenia	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
137	Georgia	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
138	Azerbaijan	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
139	Russia Asian people groups (Tatar, Chechen, Dagestan, etc.)	Direct Map (geographic category)	Other Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
140	Kazakhstan	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
141	Uzbekistan	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
142	Tadjikistan	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
143	Kyrgyzstan	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
144	Turkmenistan	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
145	Other Central Asia (includes Afghanistan)	Direct Map (geographic category)	Other Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
150	Sri Lanka	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
151	India	Direct Map (geographic category)	Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian
152	Other South Asia (includes Pakistan, Bangladesh, Himalayan countries)	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
153	Burma/Myanmar	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
154	Laos/Hmong/Iu Mienh	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
155	Cambodia/Kampuchea	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
156	Indonesia/Bali/Java	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
157	Malaysia	Indirect Map	Census	Direct Map	Other Pacific Islander	Indirect Map	Census

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
158	Malay	Direct Map	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
159	Singapore	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Indirect Map ¹	Census
160	Thailand	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
161	Thai	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
162	Mongolia	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
163	Tibet	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
164	Other East Asia	Direct Map (geographic category)	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
165	Djibouti	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
166	Sudan	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
167	Other Eastern Africa	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
168	South Africa	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
169	Namibia	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
170	Zimbabwe	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
171	Zambia	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
172	Botswana	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
173	Angola	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
174	Mozambique	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
175	Mauritius	Indirect Map	Census	Direct Map	Other Pacific Islander	Indirect Map	Census

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
176	Other Southern Africa	Direct Map (geographic category)	Black/ African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
177	Cape Verde	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
178	Sao Tome	Direct Map (geographic category)	White and Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
179	Mauritania	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
180	Mali	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
181	Niger	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
182	Other Western Africa	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
183	Chad	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
184	Other Central Africa	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian
185	African/Africa	Direct Map (geographic category)	Black/African American	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
186	Australia	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
187	New Zealand	Indirect Map ¹	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
188	Fiji	Direct Map (geographic category)	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Asian
189	Nauru	Direct Map (geographic category)	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Asian
190	Samoa	Direct Map (geographic category)	Samoa	Direct Map	Samoa	Direct Map	Samoa
191	Samoa	Direct Map	Samoa	Direct Map	Samoa	Direct Map	Samoa

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
192	Other Oceania	Direct Map (geographic category)	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Asian
193	European (nonspecific)	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
194	Cape Verde & Portuguese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
195	Cape Verde & Mexican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
196	Cape Verde & European (not Spanish)	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
201	Biracial (nonspecific)	Informative for Imputation	Restricted Imputation	Direct Map	Other Pacific Islander	Direct Map	Other Asian
202	White & Black/ African American	Direct Map	White & Black/ African American	Direct Map	White & Black/ African American	Direct Map	White & Black/ African American
203	White & American Indian/Alaska Native (including mestizo)	Direct Map	White & American Indian/Alaska Native	Direct Map	White & American Indian/Alaska Native	Direct Map	White & American Indian/Alaska Native
204	White & Native Hawaiian	Direct Map	White & Native Hawaiian	Direct Map	White & Native Hawaiian	Direct Map	White & Native Hawaiian
205	White & Other Pacific Islander	Direct Map	White & Other Pacific Islander	Direct Map	White & Other Pacific Islander	Direct Map	White & Other Pacific Islander
206	White & Asian Indian	Direct Map	White & Asian Indian	Direct Map	White & Asian Indian	Direct Map	White & Asian Indian
207	White & Chinese	Direct Map	White & Chinese	Direct Map	White & Chinese	Direct Map	White & Chinese
208	White & Filipino	Direct Map	White & Filipino	Direct Map	White & Filipino	Direct Map	White & Filipino
209	White & Japanese	Direct Map	White & Japanese	Direct Map	White & Japanese	Direct Map	White & Japanese
210	White & Korean	Direct Map	White & Korean	Direct Map	White & Korean	Direct Map	White & Korean

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
211	White & Vietnamese	Direct Map	White & Vietnamese	Direct Map	White & Vietnamese	Direct Map	White & Vietnamese
212	White & Other Asian	Direct Map	White & Other Asian	Direct Map	White & Other Asian	Direct Map	White & Other Asian
213	White & Asian (nonspecific)	Direct Map	White & Asian (nonspecific)	Direct Map	White & Asian (nonspecific)	Direct Map	White & Asian (nonspecific)
214	White & Indian (Asian or American unclear)	Direct Map	White & American Indian/Alaska Native	Direct Map	White & Asian Indian	Direct Map	White & Asian Indian
215	White & Guamanian/Chamorro	Direct Map	White & Guamanian/Chamorro	Direct Map	White & Guamanian/Chamorro	Direct Map	White & Guamanian/Chamorro
223	Black/African American & American Indian/ Alaska Native	Direct Map	Black/African American & American Indian/ Alaska Native	Direct Map	Black/African American & American Indian/ Alaska Native	Direct Map	Black/African American & American Indian/ Alaska Native
224	Black/African American & Native Hawaiian	Direct Map	Black/African American & Native Hawaiian	Direct Map	Black/African American & Native Hawaiian	Direct Map	Black/African American & Native Hawaiian
225	Black/African American & Other Pacific Islander	Direct Map	Black/African American & Other Pacific Islander	Direct Map	Black/African American & Other Pacific Islander	Direct Map	Black/African American & Other Pacific Islander
226	Black/African American & Asian Indian	Direct Map	Black/African American & Asian Indian	Direct Map	Black/African American & Asian Indian	Direct Map	Black/African American & Asian Indian
227	Black/African American & Chinese	Direct Map	Black/African American & Chinese	Direct Map	Black/African American & Chinese	Direct Map	Black/African American & Chinese
228	Black/African American & Filipino	Direct Map	Black/African American & Filipino	Direct Map	Black/African American & Filipino	Direct Map	Black/African American & Filipino

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
229	Black/African American & Japanese	Direct Map	Black/African American & Japanese	Direct Map	Black/African American & Japanese	Direct Map	Black/African American & Japanese
230	Black/African American & Korean	Direct Map	Black/African American & Korean	Direct Map	Black/African American & Korean	Direct Map	Black/African American & Korean
231	Black/African American & Vietnamese	Direct Map	Black/African American & Vietnamese	Direct Map	Black/African American & Vietnamese	Direct Map	Black/African American & Vietnamese
232	Black/African American & Other Asian	Direct Map	Black/African American & Other Asian	Direct Map	Black/African American & Other Asian	Direct Map	Black/African American & Other Asian
233	Black/African American & Asian (nonspecific)	Direct Map	Black/African American & Asian (nonspecific)	Direct Map	Black/African American & Asian (nonspecific)	Direct Map	Black/African American & Asian (nonspecific)
234	Black/African American & Indian (Asian or American unclear)	Direct Map	Black/African American & American Indian/Alaska Native	Direct Map	Black/African American & Asian Indian	Direct Map	Black/African American & Asian Indian
244	American Indian/Alaska Native & Native Hawaiian	Direct Map	American Indian/Alaska Native & Native Hawaiian	Direct Map	American Indian/Alaska Native & Native Hawaiian	Direct Map	American Indian/Alaska Native & Native Hawaiian
245	American Indian/Alaska Native & Other Pacific Islander	Direct Map	American Indian/Alaska Native & Other Pacific Islander	Direct Map	American Indian/Alaska Native & Other Pacific Islander	Direct Map	American Indian/Alaska Native & Other Pacific Islander
246	American Indian/Alaska Native & Asian Indian	Direct Map	American Indian/Alaska Native & Asian Indian	Direct Map	American Indian/Alaska Native & Asian Indian	Direct Map	American Indian/Alaska Native & Asian Indian
247	American Indian/Alaska Native & Chinese	Direct Map	American Indian/Alaska Native & Chinese	Direct Map	American Indian/Alaska Native & Chinese	Direct Map	American Indian/Alaska Native & Chinese

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
248	American Indian/ Alaska Native & Filipino	Direct Map	American Indian/ Alaska Native & Filipino	Direct Map	American Indian/ Alaska Native & Filipino	Direct Map	American Indian/ Alaska Native & Filipino
249	American Indian/ Alaska Native & Japanese	Direct Map	American Indian/ Alaska Native & Japanese	Direct Map	American Indian/ Alaska Native & Japanese	Direct Map	American Indian/ Alaska Native & Japanese
250	American Indian/ Alaska Native & Korean	Direct Map	American Indian/ Alaska Native & Korean	Direct Map	American Indian/ Alaska Native & Korean	Direct Map	American Indian/ Alaska Native & Korean
251	American Indian/ Alaska Native & Vietnamese	Direct Map	American Indian/ Alaska Native & Vietnamese	Direct Map	American Indian/ Alaska Native & Vietnamese	Direct Map	American Indian/ Alaska Native & Vietnamese
252	American Indian/ Alaska Native & Other Asian	Direct Map	American Indian/ Alaska Native & Other Asian	Direct Map	American Indian/ Alaska Native & Other Asian	Direct Map	American Indian/ Alaska Native & Other Asian
253	American Indian/ Alaska Native & Asian (nonspecific)	Direct Map	American Indian/ Alaska Native & Asian (nonspecific)	Direct Map	American Indian/ Alaska Native & Asian (nonspecific)	Direct Map	American Indian/ Alaska Native & Asian (nonspecific)
265	Native Hawaiian & Other Pacific Islander	Direct Map	Native Hawaiian & Other Pacific Islander	Direct Map	Native Hawaiian & Other Pacific Islander	Direct Map	Native Hawaiian & Other Pacific Islander
266	Native Hawaiian & Asian Indian	Direct Map	Native Hawaiian & Asian Indian	Direct Map	Native Hawaiian & Asian Indian	Direct Map	Native Hawaiian & Asian Indian
267	Native Hawaiian & Chinese	Direct Map	Native Hawaiian & Chinese	Direct Map	Native Hawaiian & Chinese	Direct Map	Native Hawaiian & Chinese
268	Native Hawaiian & Filipino	Direct Map	Native Hawaiian & Filipino	Direct Map	Native Hawaiian & Filipino	Direct Map	Native Hawaiian & Filipino
269	Native Hawaiian & Japanese	Direct Map	Native Hawaiian & Japanese	Direct Map	Native Hawaiian & Japanese	Direct Map	Native Hawaiian & Japanese

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
270	Native Hawaiian & Korean	Direct Map	Native Hawaiian & Korean	Direct Map	Native Hawaiian & Korean	Direct Map	Native Hawaiian & Korean
271	Native Hawaiian & Vietnamese	Direct Map	Native Hawaiian & Vietnamese	Direct Map	Native Hawaiian & Vietnamese	Direct Map	Native Hawaiian & Vietnamese
272	Native Hawaiian & Other Asian	Direct Map	Native Hawaiian & Other Asian	Direct Map	Native Hawaiian & Other Asian	Direct Map	Native Hawaiian & Other Asian
273	Native Hawaiian & Asian (nonspecific)	Direct Map	Native Hawaiian & Asian (nonspecific)	Direct Map	Native Hawaiian & Asian (nonspecific)	Direct Map	Native Hawaiian & Asian (nonspecific)
274	Native Hawaiian, Samoan, & Other Pacific Islander	Direct Map	Native Hawaiian, Samoan, & Other Pacific Islander	Direct Map	Native Hawaiian, Samoan, & Other Pacific Islander	Direct Map	Native Hawaiian, Samoan, & Other Pacific Islander
286	Other Pacific Islander & Asian Indian	Direct Map	Other Pacific Islander & Asian Indian	Direct Map	Other Pacific Islander & Asian Indian	Direct Map	Other Pacific Islander & Asian Indian
287	Other Pacific Islander & Chinese	Direct Map	Other Pacific Islander & Chinese	Direct Map	Other Pacific Islander & Chinese	Direct Map	Other Pacific Islander & Chinese
288	Other Pacific Islander & Filipino	Direct Map	Other Pacific Islander & Filipino	Direct Map	Other Pacific Islander & Filipino	Direct Map	Other Pacific Islander & Filipino
289	Other Pacific Islander & Japanese	Direct Map	Other Pacific Islander & Japanese	Direct Map	Other Pacific Islander & Japanese	Direct Map	Other Pacific Islander & Japanese
290	Other Pacific Islander & Korean	Direct Map	Other Pacific Islander & Korean	Direct Map	Other Pacific Islander & Korean	Direct Map	Other Pacific Islander & Korean
291	Other Pacific Islander & Vietnamese	Direct Map	Other Pacific Islander & Vietnamese	Direct Map	Other Pacific Islander & Vietnamese	Direct Map	Other Pacific Islander & Vietnamese
292	Other Pacific Islander & Other Asian	Direct Map	Other Pacific Islander & Other Asian	Direct Map	Other Pacific Islander & Other Asian	Direct Map	Other Pacific Islander & Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
293	Other Pacific Islander & Asian (nonspecific)	Direct Map	Other Pacific Islander & Asian (nonspecific)	Direct Map	Other Pacific Islander & Asian (nonspecific)	Direct Map	Other Pacific Islander & Asian (nonspecific)
294	Other Pacific Islander & Samoan	Direct Map	Other Pacific Islander & Samoan	Direct Map	Other Pacific Islander & Samoan	Direct Map	Other Pacific Islander & Samoan
307	Asian Indian & Chinese	Direct Map	Asian Indian & Chinese	Direct Map	Asian Indian & Chinese	Direct Map	Asian Indian & Chinese
308	Asian Indian & Filipino	Direct Map	Asian Indian & Filipino	Direct Map	Asian Indian & Filipino	Direct Map	Asian Indian & Filipino
309	Asian Indian & Japanese	Direct Map	Asian Indian & Japanese	Direct Map	Asian Indian & Japanese	Direct Map	Asian Indian & Japanese
310	Asian Indian & Korean	Direct Map	Asian Indian & Korean	Direct Map	Asian Indian & Korean	Direct Map	Asian Indian & Korean
311	Asian Indian & Vietnamese	Direct Map	Asian Indian & Vietnamese	Direct Map	Asian Indian & Vietnamese	Direct Map	Asian Indian & Vietnamese
312	Asian Indian & Other Asian	Direct Map	Asian Indian & Other Asian	Direct Map	Asian Indian & Other Asian	Direct Map	Asian Indian & Other Asian
328	Chinese & Filipino	Direct Map	Chinese & Filipino	Direct Map	Chinese & Filipino	Direct Map	Chinese & Filipino
329	Chinese & Japanese	Direct Map	Chinese & Japanese	Direct Map	Chinese & Japanese	Direct Map	Chinese & Japanese
330	Chinese & Korean	Direct Map	Chinese & Korean	Direct Map	Chinese & Korean	Direct Map	Chinese & Korean
331	Chinese & Vietnamese	Direct Map	Chinese & Vietnamese	Direct Map	Chinese & Vietnamese	Direct Map	Chinese & Vietnamese
332	Chinese & Other Asian	Direct Map	Chinese & Other Asian	Direct Map	Chinese & Other Asian	Direct Map	Chinese & Other Asian
349	Filipino & Japanese	Direct Map	Filipino & Japanese	Direct Map	Filipino & Japanese	Direct Map	Filipino & Japanese
350	Filipino & Korean	Direct Map	Filipino & Korean	Direct Map	Filipino & Korean	Direct Map	Filipino & Korean

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
351	Filipino & Vietnamese	Direct Map	Filipino & Vietnamese	Direct Map	Filipino & Vietnamese	Direct Map	Filipino & Vietnamese
352	Filipino & Other Asian	Direct Map	Filipino & Other Asian	Direct Map	Filipino & Other Asian	Direct Map	Filipino & Other Asian
360	Japanese & Korean	Direct Map	Japanese & Korean	Direct Map	Japanese & Korean	Direct Map	Japanese & Korean
361	Japanese & Vietnamese	Direct Map	Japanese & Vietnamese	Direct Map	Japanese & Vietnamese	Direct Map	Japanese & Vietnamese
362	Japanese & Other Asian	Direct Map	Japanese & Other Asian	Direct Map	Japanese & Other Asian	Direct Map	Japanese & Other Asian
371	Korean & Vietnamese	Direct Map	Korean & Vietnamese	Direct Map	Korean & Vietnamese	Direct Map	Korean & Vietnamese
372	Korean & Other Asian	Direct Map	Korean & Other Asian	Direct Map	Korean & Other Asian	Direct Map	Korean & Other Asian
382	Vietnamese & Other Asian	Direct Map	Vietnamese & Other Asian	Direct Map	Vietnamese & Other Asian	Direct Map	Vietnamese & Other Asian
383	Indian (Asian or American unclear) & Native Hawaiian	Direct Map	American Indian/ Alaska Native & Native Hawaiian	Direct Map	Asian Indian & Native Hawaiian	Direct Map	Asian Indian & Native Hawaiian
384	Indian (Asian or American unclear) & Other Pacific Islander	Direct Map	American Indian/ Alaska Native & Other Pacific Islander	Direct Map	Asian Indian & Other Pacific Islander	Direct Map	Asian Indian & Other Pacific Islander
385	Indian (Asian or American unclear) & Chinese	Direct Map	American Indian/ Alaska Native & Chinese	Direct Map	Asian Indian & Chinese	Direct Map	Asian Indian & Chinese
386	Indian (Asian or American unclear) & Filipino	Direct Map	American Indian/ Alaska Native & Filipino	Direct Map	Asian Indian & Filipino	Direct Map	Asian Indian & Filipino

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
387	Indian (Asian or American unclear) & Japanese	Direct Map	American Indian/Alaska Native & Japanese	Direct Map	Asian Indian & Japanese	Direct Map	Asian Indian & Japanese
388	Indian (Asian or American unclear) & Korean	Direct Map	American Indian/Alaska Native & Korean	Direct Map	Asian Indian & Korean	Direct Map	Asian Indian & Korean
389	Indian (Asian or American unclear) & Vietnamese	Direct Map	American Indian/Alaska Native & Vietnamese	Direct Map	Asian Indian & Vietnamese	Direct Map	Asian Indian & Vietnamese
390	Indian (Asian or American unclear) & Other Asian	Direct Map	American Indian/Alaska Native & Other Asian	Direct Map	Asian Indian & Other Asian	Direct Map	Asian Indian & Other Asian
391	Indian (Asian or American unclear) & Jamaican	Direct Map	American Indian/Alaska Native & Jamaican	Direct Map	Asian Indian & Jamaican	Direct Map	Asian Indian & Jamaican
401	White, Black/African American, American Indian/Alaska Native	Direct Map	White, Black/African American, American Indian/Alaska Native	Direct Map	White, Black/African American, American Indian/Alaska Native	Direct Map	White, Black/African American, American Indian/Alaska Native
402	White, Black/African American, Native Hawaiian	Direct Map	White, Black/African American, Native Hawaiian	Direct Map	White, Black/African American, Native Hawaiian	Direct Map	White, Black/African American, Native Hawaiian
403	White, Black/African American, Other Pacific Islander	Direct Map	White, Black/African American, Other Pacific Islander	Direct Map	White, Black/African American, Other Pacific Islander	Direct Map	White, Black/African American, Other Pacific Islander
404	White, Black/African American, Asian Indian	Direct Map	White, Black/African American, Asian Indian	Direct Map	White, Black/African American, Asian Indian	Direct Map	White, Black/African American, Asian Indian
405	White, Black/African American, Chinese	Direct Map	White, Black/African American, Chinese	Direct Map	White, Black/African American, Chinese	Direct Map	White, Black/African American, Chinese

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
406	White, Black/African American, Filipino	Direct Map	White, Black/African American, Filipino	Direct Map	White, Black/African American, Filipino	Direct Map	White, Black/African American, Filipino
407	White, Black/African American, Japanese	Direct Map	White, Black/African American, Japanese	Direct Map	White, Black/African American, Japanese	Direct Map	White, Black/African American, Japanese
408	White, Black/African American, Korean	Direct Map	White, Black/African American, Korean	Direct Map	White, Black/African American, Korean	Direct Map	White, Black/African American, Korean
409	White, Black/African American, Vietnamese	Direct Map	White, Black/African American, Vietnamese	Direct Map	White, Black/African American, Vietnamese	Direct Map	White, Black/African American, Vietnamese
410	White, Black/African American, Other Asian	Direct Map	White, Black/African American, Other Asian	Direct Map	White, Black/African American, Other Asian	Direct Map	White, Black/African American, Other Asian
411	White, Black/African American, Asian (nonspecific)	Direct Map	White, Black/African American, Asian (nonspecific)	Direct Map	White, Black/African American, Asian (nonspecific)	Direct Map	White, Black/African American, Asian (nonspecific)
412	White, American Indian/Alaska Native, Native Hawaiian	Direct Map	White, American Indian/Alaska Native, Native Hawaiian	Direct Map	White, American Indian/Alaska Native, Native Hawaiian	Direct Map	White, American Indian/Alaska Native, Native Hawaiian
413	White, American Indian/Alaska Native, Other Pacific Islander	Direct Map	White, American Indian/Alaska Native, Other Pacific Islander	Direct Map	White, American Indian/Alaska Native, Other Pacific Islander	Direct Map	White, American Indian/Alaska Native, Other Pacific Islander
414	White, American Indian/Alaska Native, Asian Indian	Direct Map	White, American Indian/Alaska Native, Asian Indian	Direct Map	White, American Indian/Alaska Native, Asian Indian	Direct Map	White, American Indian/Alaska Native, Asian Indian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
415	White, American Indian/Alaska Native, Chinese	Direct Map	White, American Indian/Alaska Native, Chinese	Direct Map	White, American Indian/Alaska Native, Chinese	Direct Map	White, American Indian/Alaska Native, Chinese
416	White, American Indian/Alaska Native, Filipino	Direct Map	White, American Indian/Alaska Native, Filipino	Direct Map	White, American Indian/Alaska Native, Filipino	Direct Map	White, American Indian/Alaska Native, Filipino
417	White, American Indian/Alaska Native, Japanese	Direct Map	White, American Indian/Alaska Native, Japanese	Direct Map	White, American Indian/Alaska Native, Japanese	Direct Map	White, American Indian/Alaska Native, Japanese
418	White, American Indian/Alaska Native, Korean	Direct Map	White, American Indian/Alaska Native, Korean	Direct Map	White, American Indian/Alaska Native, Korean	Direct Map	White, American Indian/Alaska Native, Korean
419	White, American Indian/Alaska Native, Vietnamese	Direct Map	White, American Indian/Alaska Native, Vietnamese	Direct Map	White, American Indian/Alaska Native, Vietnamese	Direct Map	White, American Indian/Alaska Native, Vietnamese
420	White, American Indian/Alaska Native, Other Asian	Direct Map	White, American Indian/Alaska Native, Other Asian	Direct Map	White, American Indian/Alaska Native, Other Asian	Direct Map	White, American Indian/Alaska Native, Other Asian
421	White, American Indian/Alaska Native, Asian (nonspecific)	Direct Map	White, American Indian/Alaska Native, Asian (nonspecific)	Direct Map	White, American Indian/Alaska Native, Asian (nonspecific)	Direct Map	White, American Indian/Alaska Native, Asian (nonspecific)
422	White, Black/African American, Indian (Asian or American unclear)	Direct Map	White, Black/African American, American Indian/Alaska Native	Direct Map	White, Black/African American, Asian Indian	Direct Map	White, Black/African American, Asian Indian
423	White, Native Hawaiian, Indian (Asian or American unclear)	Direct Map	White, Native Hawaiian, American Indian/Alaska Native	Direct Map	White, Native Hawaiian, Asian Indian	Direct Map	White, Native Hawaiian, Asian Indian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
424	White, Other Pacific Islander, Indian (Asian or American unclear)	Direct Map	White, Other Pacific Islander, American Indian/Alaska Native	Direct Map	White, Other Pacific Islander, Asian Indian	Direct Map	White, Other Pacific Islander, Asian Indian
450	Belize and White	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
600	Mexican & Guatemalan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
601	Mexican & El Salvadoran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
602	Mexican & Honduran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
603	Mexican & Nicaraguan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
604	Mexican & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
605	Mexican & Panamanian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
606	Mexican & Colombian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
607	Mexican & Venezuelan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
608	Mexican & Ecuadorian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
609	Mexican & Peruvian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
610	Mexican & Bolivian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
611	Mexican & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
612	Mexican & Argentine	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
613	Mexican & Paraguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
614	Mexican & Uruguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
615	Mexican & Brazilian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
616	Puerto Rican & Guatemalan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
617	Puerto Rican & El Salvadoran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
618	Puerto Rican & Honduran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
619	Puerto Rican & Nicaraguan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
620	Puerto Rican & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
621	Puerto Rican & Panamanian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
622	Puerto Rican & Colombian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
623	Puerto Rican & Venezuelan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
624	Puerto Rican & Ecuadorian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
625	Puerto Rican & Peruvian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
626	Puerto Rican & Bolivian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
627	Puerto Rican & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
628	Puerto Rican & Argentine	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
629	Puerto Rican & Paraguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
630	Puerto Rican & Uruguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
631	Puerto Rican & Brazilian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
632	Cuban & Guatemalan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
633	Cuban & El Salvadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
634	Cuban & Honduran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
635	Cuban & Nicaraguan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
636	Cuban & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
637	Cuban & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
638	Cuban & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
639	Cuban & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
640	Cuban & Ecuadorian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
641	Cuban & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
642	Cuban & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
643	Cuban & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
644	Cuban & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
645	Cuban & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
646	Cuban & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
647	Cuban & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
648	Dominican & Guatemalan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
649	Dominican & El Salvadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
650	Dominican & Honduran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
651	Dominican & Nicaraguan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
652	Dominican & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
653	Dominican & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
654	Dominican & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
655	Dominican & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
656	Dominican & Ecuadorian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
657	Dominican & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
658	Dominican & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
659	Dominican & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
660	Dominican & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
661	Dominican & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
662	Dominican & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
663	Dominican & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
664	Spanish (from Spain) & Guatemalan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
665	Spanish (from Spain) & El Salvadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
666	Spanish (from Spain) & Honduran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
667	Spanish (from Spain) & Nicaraguan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
668	Spanish (from Spain) & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
669	Spanish (from Spain) & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
670	Spanish (from Spain) & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
671	Spanish (from Spain) & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
672	Spanish (from Spain) & Ecuadorian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
673	Spanish (from Spain) & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
674	Spanish (from Spain) & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
675	Spanish (from Spain) & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
676	Spanish (from Spain) & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
677	Spanish (from Spain) & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
678	Spanish (from Spain) & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
679	Spanish (from Spain) & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
680	Colombian & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
681	Colombian & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
682	Colombian & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
683	Colombian & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
684	Colombian & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
685	Colombian & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
686	Colombian & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
687	Colombian & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
688	Colombian & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
689	Venezuelan & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
690	Venezuelan & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
691	Venezuelan & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
692	Venezuelan & Chilean	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
693	Venezuelan & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
694	Venezuelan & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
695	Venezuelan & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
696	Venezuelan & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
697	Ecuadoran & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
698	Ecuadoran & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
699	Ecuadoran & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
700	Ecuadoran & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
701	Ecuadoran & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
702	Ecuadoran & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
703	Ecuadoran & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
704	Peruvian & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
705	Peruvian & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
706	Peruvian & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
707	Peruvian & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
708	Peruvian & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
709	Peruvian & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
710	Bolivian & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
711	Bolivian & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
712	Bolivian & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
713	Bolivian & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
714	Bolivian & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
715	Chilean & Argentine	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
716	Chilean & Paraguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
717	Chilean & Uruguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
718	Chilean & Brazilian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
719	Argentine & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
720	Argentine & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
721	Argentine & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
722	Paraguayan & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
723	Paraguayan & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
724	Uruguayan & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
725	Guatemalan & Colombian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
726	El Salvadoran & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
727	Honduran & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 **Descriptions of Race Codes and the Groups to Which They Mapped (continued)**

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
728	Nicaraguan & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
729	Costa Rican & Colombian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
730	Panamanian & Colombian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
731	Guatemalan & Venezuelan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
732	El Salvadoran & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
733	Honduran & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
734	Nicaraguan & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
735	Costa Rican & Venezuelan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
736	Panamanian & Venezuelan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
737	Guatemalan & Ecuadoran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
738	El Salvadoran & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
739	Honduran & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
740	Nicaraguan & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
741	Costa Rican & Ecuadoran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
742	Panamanian & Ecuadoran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
743	Guatemalan & Peruvian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
744	El Salvadoran & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
745	Honduran & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
746	Nicaraguan & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
747	Costa Rican & Peruvian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
748	Panamanian & Peruvian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
749	Guatemalan & Bolivian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
750	El Salvadoran & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
751	Honduran & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
752	Nicaraguan & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
753	Costa Rican & Bolivian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
754	Panamanian & Bolivian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
755	Guatemalan & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
756	El Salvadoran & Chilean	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
757	Honduran & Chilean	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
758	Nicaraguan & Chilean	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
759	Costa Rican & Chilean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
760	Panamanian & Chilean	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
761	Guatemalan & Argentine	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
762	El Salvadoran & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
763	Honduran & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
764	Nicaraguan & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
765	Costa Rican & Argentine	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
766	Panamanian & Argentine	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
767	Guatemalan & Paraguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
768	El Salvadoran & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
769	Honduran & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
770	Nicaraguan & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
771	Costa Rican & Paraguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
772	Panamanian & Paraguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
773	Guatemalan & Uruguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
774	El Salvadoran & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
775	Honduran & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
776	Nicaraguan & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
777	Costa Rican & Uruguayan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
778	Panamanian & Uruguayan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
779	Guatemalan & Brazilian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
780	El Salvadoran & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
781	Honduran & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
782	Nicaraguan & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
783	Costa Rican & Brazilian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
784	Panamanian & Brazilian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
785	Guatemalan & El Salvadoran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
786	Guatemalan & Honduran	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
787	Guatemalan & Nicaraguan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
788	Guatemalan & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
789	Guatemalan & Panamanian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
790	El Salvadoran & Honduran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
791	El Salvadoran & Nicaraguan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
792	El Salvadoran & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
793	El Salvadoran & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
794	Honduran & Nicaraguan	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
795	Honduran & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
796	Honduran & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
797	Nicaraguan & Costa Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
798	Nicaraguan & Panamanian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
799	Costa Rican & Panamanian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
800	Mexican & Jamaican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
801	Puerto Rican & Jamaican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
802	Cuban & Jamaican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
803	Dominican & Jamaican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
804	Spanish (from Spain) & Jamaican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
805	Mexican & European (not Spanish)	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
806	Puerto Rican & European (not Spanish)	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
807	Cuban & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
808	Dominican & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
809	Spanish (from Spain) & Other European	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
810	Trinidadian & Mexican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
811	Trinidadian & Puerto Rican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
812	Trinidadian & Cuban	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
813	Trinidadian & Dominican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
814	Trinidadian & Spanish (from Spain)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
815	Guatemalan & European (not Spanish)	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
816	El Salvador & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
817	Honduran & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
818	Nicaraguan & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
819	Costa Rican & European (not Spanish)	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
820	Panamanian & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
821	Colombian & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
822	Venezuelan & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
823	Ecuadoran & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
824	Peruvian & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
825	Bolivian & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
826	Chilean & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
827	Argentine & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
828	Paraguayan & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
829	Uruguayan & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
830	Brazilian & European (not Spanish)	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
831	(part) Mexican, ½ (part) White	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
832	(part) Mexican, ½ (part) Black/African American	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
833	(part) Mexican, ½ (part) American Indian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
834	(part) Mexican, ½ (part) Hawaiian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
835	(part) Mexican, ½ (part) Other Pacific Islander	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
836	(part) Mexican, ½ (part) Asian Indian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
837	(part) Mexican, ½ (part) Chinese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
838	(part) Mexican, ½ (part) Filipino	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
839	(part) Mexican, ½ (part) Japanese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
840	(part) Mexican, ½ (part) Korean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
841	(part) Mexican, ½ (part) Vietnamese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
842	(part) Mexican, ½ (part) Other Asian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
843	(part) Puerto Rican, ½ (part) White	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
844	(part) Puerto Rican, ½ (part) Black/African American	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
845	(part) Puerto Rican, ½ (part) American Indian/Alaska Native	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
846	(part) Puerto Rican, ½ (part) Native Hawaiian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
847	(part) Puerto Rican, ½ (part) Other Pacific Islander	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
848	(part) Puerto Rican, ½ (part) Asian Indian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
849	(part) Puerto Rican, ½ (part) Chinese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
850	(part) Puerto Rican, ½ (part) Filipino	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
851	(part) Puerto Rican, ½ (part) Japanese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
852	(part) Puerto Rican, ½ (part) Korean	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
853	(part) Puerto Rican, ½ (part) Vietnamese	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
854	(part) Puerto Rican, ½ (part) Other Asian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
855*	(part) Hispanic/Latino, ½ (part) White	Direct Map	White	Direct Map	White	Direct Map	White
856*	(part) Hispanic/Latino, ½ (part) Black/African American	Direct Map	Black/African American	Direct Map	Black/African American	Direct Map	Black/African American
857*	(part) Hispanic/Latino, ½ (part) American Indian/Alaska Native	Direct Map	American Indian/Alaska Native	Direct Map	American Indian/Alaska Native	Direct Map	American Indian/Alaska Native
858*	(part) Hispanic/Latino, ½ (part) Native Hawaiian	Direct Map	Native Hawaiian	Direct Map	Native Hawaiian	Direct Map	Native Hawaiian
859*	(part) Hispanic/Latino, ½ (part) Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander
860*	(part) Hispanic/Latino, ½ (part) Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
861*	(part) Hispanic/Latino, ½ (part) Chinese	Direct Map	Chinese	Direct Map	Chinese	Direct Map	Chinese
862*	(part) Hispanic/Latino, ½ (part) Filipino	Direct Map	Filipino	Direct Map	Filipino	Direct Map	Filipino
863*	(part) Hispanic/Latino, ½ (part) Japanese	Direct Map	Japanese	Direct Map	Japanese	Direct Map	Japanese
864*	(part) Hispanic/Latino, ½ (part) Korean	Direct Map	Korean	Direct Map	Korean	Direct Map	Korean
865*	(part) Hispanic/Latino, ½ (part) Vietnamese	Direct Map	Vietnamese	Direct Map	Vietnamese	Direct Map	Vietnamese
866*	(part) Hispanic/Latino, ½ (part) Other Asian	Direct Map	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
869	Haitian & Jamaican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
870	Haitian & Dominican	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
871	Honduran & Haitian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
872	Guatemalan & Iranian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
873	Panamanian & Jamaican	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
874	Cuban & Thai	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
875	Venezuelan & Trinidadian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
876	Puerto Rican & Arab	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
877	Puerto Rican & Virgin Islands	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
878	Mexican & Samoan	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
879	Salvadoran & Egyptian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
880	Costa Rican & Haitian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
881	Mexican & Iranian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
882	Spanish & Barbadian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
883	Peruvian & Other Middle Eastern	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
884	Puerto Rican & African	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
885	Jamaican & Egyptian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
886	Argentine & Turkish	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
887	Mexican & Egyptian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
888	Guatemalan & Canadian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
889	Haitian & European	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
890	Argentine, Cuban, & Spanish	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
891	Mexican, Cuban, & European	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
892	Mexican, Puerto Rican, & European	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
893	Haitian & Trinidadian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
894	Belizean & Honduran	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
895	Trinidadian & European	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
896	Puerto Rican & Haitian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
897	Chinese & Malaysian	Indirect Map	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
898	Cuban, Chinese, & White	Indirect Map	Triple Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
900*	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain or Dominican Republic)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
901*	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
902*	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
903	Central/South American (no country)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
904	Nonwhite (nonspecific/ brown)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
905*	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
906*	Mezclado, Mezclada (Hispanic/Latino mixed)	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
907	Mixed	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
908	Olive	Direct Map	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
909	Creole	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
910	Arab	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
911	Jewish	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
912	Kurd	Direct Map (geographic category)	Other Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
913	Chaldean/ Caldanian/ Assyrian	Direct Map (geographic category)	Other Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
914	Romany/Gypsy	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
915	Central/South American & West Indies	Indirect Map	Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
916	Central/South American & Mexican	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
917	Central/South American & Puerto Rican	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
918	Central/South American & Cuban	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
919	Central/South American & Dominican	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
920	Central/South American & Spanish	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
921	Arab/Asian	Direct Map	White & Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
922	Arab/European	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
923	Arab/African	Direct Map (geographic category)	White & Black/African American	Direct Map	Other Pacific Islander	Direct Map	Other Asian & Black/African American
924	Arab/Chaldean	Direct Map (geographic category)	White & Asian	Direct Map	Other Pacific Islander	Direct Map	Other Asian
925	European & Asian Indian	Direct Map (geographic category)	White & Asian Indian	Direct Map	White & Asian Indian	Direct Map	White & Asian Indian
926	West Indies & Belize	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
927	West Indies & Cape Verde	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
928	Arab & Cape Verde	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
929	Aryan	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Asian Indian
930	Turkish & Lebanese	Direct Map (geographic category)	White	Direct Map	Other Pacific Islander	Direct Map	Other Asian
934	Dominican & Asian Indian	Indirect Map ¹	Double Census	Direct Map	Asian Indian	Direct Map	Asian Indian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
937	Cuban & Indian (Asian or American unclear)	Indirect Map	Double Census	Direct Map	Asian Indian	Direct & Indirect Map	Asian Indian
938	Cuban & American Indian	Indirect Map ¹	Double Census	Direct Map	Asian Indian	Direct Map	Asian Indian
939	Cuban & Asian Indian	Indirect Map ¹	Double Census	Direct Map	Asian Indian	Direct Map ¹	Asian Indian
940	Thai & Malaysian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
941	Jamaican & American Indian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
942	Jamaican & Asian Indian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Indirect Map ¹	Double Census
943	Jamaican & Colombian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
944	White & Malaysian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
945	Grenada & Brazil	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
946	Canadian & Haitian	Indirect Map ¹	Double Census	Direct Map	Other Pacific Islander	Direct Map	Other Asian
950	White, Black, and Something Else	Direct Map	White & Black	Direct Map	White, Black & Other Pacific	Direct Map	White, Black & Other Asian
951	White and Something Else	Direct Map	White	Direct Map	White & Other Pacific	Direct Map	White & Other Asian
952	Black/African American and Something Else	Direct Map	Black/African American	Direct Map	Black/African American & Other Pacific	Direct Map	Black/African American & Other Asian
953	American Indian/Alaska Native and Something Else	Direct Map	American Indian/Alaska Native	Direct Map	American Indian/Alaska Native & Other Pacific	Direct Map	American Indian/Alaska Native & Other Asian
954	Native Hawaiian and Something Else	Direct Map	Native Hawaiian	Direct Map	Native Hawaiian & Other Pacific	Direct Map	Native Hawaiian & Other Asian
955	Other Pacific Islander and Something Else	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander	Direct Map	Other Pacific Islander & Other Asian

Table C.1 Descriptions of Race Codes and the Groups to Which They Mapped (continued)

Race Code	Race Name	QD05		QD05PI		QD05Asia	
		Type ¹	Result	Type ¹	Result	Type ¹	Result
956	Asian Indian and Something Else	Direct Map	Asian Indian	Direct Map	Asian Indian	Direct Map	Asian Indian
957	Chinese and Something Else	Direct Map	Chinese	Direct Map	Chinese	Direct Map	Chinese
958	Filipino and Something Else	Direct Map	Filipino	Direct Map	Filipino	Direct Map	Filipino
959	Japanese and Something Else	Direct Map	Japanese	Direct Map	Japanese	Direct Map	Japanese
960	Korean and Something Else	Direct Map	Korean	Direct Map	Korean	Direct Map	Korean
961	Vietnamese and Something Else	Direct Map	Vietnamese	Direct Map	Vietnamese	Direct Map	Vietnamese
962	Other Asian and Something Else	Direct Map	Other Asian	Direct Map	Other Asian	Direct Map	Other Asian
963	Asian (nonspecific) and Something Else	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation	Informative for Imputation	Restricted Imputation
964	Indian (Asian or American unclear) and Something Else	Direct Map	American Indian/ Alaska Native (Multiple Race)	Direct Map	Asian Indian (Multiple Race)	Direct Map	Asian Indian (Multiple Race)
965	Brown & White	Direct Map	White & Black/ African American	Direct Map	White & Black/ African American	Direct Map	White & Black/ African American
985	Bad Data	Uninformative Code	Unrestricted Imputation	Uninformative Code	Restricted Imputation	Uninformative Code	Restricted Imputation
994	Unknown/ "Don't Know"	Uninformative Code	Unrestricted Imputation	Uninformative Code	Restricted Imputation	Uninformative Code	Restricted Imputation
997	Rather Not Say/ "Refused" ("American" or "All of Them")	Uninformative Code	Unrestricted Imputation	Uninformative Code	Restricted Imputation	Uninformative Code	Restricted Imputation

*These codes caused the Hispanic/Latino indicator to be edited to a "yes" if QD03 was missing or "no." The code that caused the Hispanic/Latino indicator to be edited to a "no" was a Hispanic/Latino code (600) and is listed in [Table C.4](#).

¹ Among the indirectly mapped codes, codes where an imputation was possible based on census information are indicated by the superscript "I." See Section C.2.1.2 for details.

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes

Race Code	Race Name	Probabilities
50	Belize	4.8% White, 11.3% American Indian/Alaska Native, 3.9% Asian Indian, 1% Asian Nonspecific, 25.9% White and Black/African American, 47% White and American Indian/Alaska Native, 6.1% Black/African American and American Indian/Alaska Native
51	Guyana	QD05: 29.3% Black/African American, 10.5% American Indian/Alaska Native, 39.8% Asian Indian, 19.9% White and Black/African American, 0.5% Unrestricted Imputation
52	Suriname	QD05: 1% White, 10% Black/African American, 2% American Indian/Alaska Native, 37% Asian Indian, 2% Chinese, 15% Other Asian, 31% White and Black/African American, 2% Unrestricted Imputation
53	Haiti	95% Black/African American, 5% White and Black/African American
54	Trinidad and Tobago	QD05: 34.2% Black/African American, 35.4% Asian Indian, 7.7% Black/African American and Asian Indian, 15.3% Impute among those with multiple races, 7.4% Unrestricted Imputation QD05 when in combination with another race: 36.7% Black/African American, 37.9% Asian Indian, 25.5% Black/African American and Asian Indian
55	Jamaica	QD05: 92.1% Black/African American, 0.8% Asian Indian, 6.1% White and Black/African American, 1% Unrestricted Imputation QD05 when in combination with another race: 92.4% Black/African American, 1.1% Asian Indian, 6.4% White and Black/African American
56	Virgin Is (St Thomas, St Croix)	QD05: 15.6% White, 76% Black/African American, 1.4% Asian Nonspecific, 2.1% Impute among those with multiple races, 4.9% Unrestricted Imputation QD05 when in combination with another race: 18.1% White, 78.5% Black/African American, 1.4% Asian Nonspecific, 0.7% White and Black/African American, 0.7% White and American Indian/Alaska Native, 0.7% Black/African American and American Indian/Alaska Native
57	Bahamas	QD05: 4.7% White, 90.6% Black/African American, 2.1% White and Black/African American, 2.6% Unrestricted Imputation
58	Barbados	2.7% White, 92.4% Black/African American, 3.1% White and Black/African American, 1.3% Asian Indian, 0.5% Unrestricted Imputation QD05 when in combination with another race: 2.9% White, 92.6% Black/African American, 3.3% White and Black/African American, 1.3% Asian Indian

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
59	Grenada	82.4% Black/African American, 2.2% Asian Indian, 13.3% White and Black/African American, 2.1% Unrestricted Imputation QD05 when in combination with another race: 83.5% Black/African American, 2.2% Asian Indian, 14.4% White and Black/African American
60	St. Lucia	QD05: 85.3% Black/African American, 2.2% Asian Indian, 5.45% White and Black/African American, 5.45% Black/African American and Asian Indian, 1.7% Unrestricted Imputation
63	Other West Indies	QD05: 80% Black/African American, 14% Asian Nonspecific, 6% Unrestricted Imputation QD05 when in combination with another race: 85.1% Black/African American, 14.9% Asian Nonspecific
64	Brazil	47.7% White, 7.6% Black/African American, 0.5% American Indian/Alaska Native, 43.1% White and Black/African American, 1.1% Asian Nonspecific
65	Canada	QD05: 86.5% White, 4.4% American Indian/Alaska Native, 4.0% Asian Indian, 5.1% Chinese
70	Mexico	QD05 when in combination with another race: 10% White, 28% American Indian/Alaska Native, 62% White and American Indian/Alaska Native
71	Puerto Rico	QD05 when in combination with another race: 75.8% White, 12.4% Black/African American, 2.8% American Indian/Alaska Native, 2.8% Native Hawaiian, 2.8% Other Pacific Islander, 1.1% White and Black/African American, 1.1% White and American Indian/Alaska Native, 1.1% Black/African American and American Indian/Alaska Native
72	Cuba	QD05 when in combination with another race: 64.1% White, 9.3% Black/African American, 26.6% White and American Indian/Alaska Native
73	Dominican Republic	QD05 when in combination with another race: 13.5% White, 15.8% Black/African American, 41.6% White and Black/African American, 29.2% White and American Indian/Alaska Native
74	Guatemala	0.2% Black/African American, 43.5% American Indian/Alaska Native, 56.0% White and American Indian/Alaska Native, 0.1% Black/African American and American Indian/Alaska Native, 0.2% Unrestricted Imputation QD05 when in combination with another race: 0.2% Black/African American, 43.6% American Indian/Alaska Native, 56.1% White and American Indian/Alaska Native, 0.1% Black/African American and American Indian/Alaska Native

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
75	Honduras	1% White, 2% Black/African American, 7% American Indian/Alaska Native, 90% White and American Indian/Alaska Native
76	El Salvador	12.7% White, 1% American Indian/Alaska Native, 86.3% White and American Indian/Alaska Native
77	Nicaragua	17% White, 9% Black/African American, 5% American Indian/Alaska Native, 69% White and American Indian/Alaska Native
78	Costa Rica	QD05: 1.1% Black/African American, 2.4% American Indian/Alaska Native, 6.7% White and Black/African American, 83.6% Imputed among White or White and American Indian, 6.2% Unrestricted Imputation QD05 when in combination with another race: 43.9% White, 1.1% Black/African American, 2.4% American Indian/Alaska Native, 8.8% White and Black/African American, 43.9% White and American Indian/Alaska Native
79	Panama	6.7% White, 9.2% Black/African American, 12.3% American Indian/Alaska Native, 6.8% White and Black/African American, 65% White and American Indian/Alaska Native
80	Colombia	22.5% White, 1.5% Black/African American, 4.3% American Indian/Alaska Native, 5.3% White and Black/African American, 65% White and American Indian/Alaska Native, 1.4% Unrestricted Imputation QD05 when in combination with another race: 23.2% White, 1.5% Black/African American, 4.3% American Indian/Alaska Native, 5.3% White and Black/African American, 65.7% White and American Indian/Alaska Native
81	Venezuela	21% White, 10% Black/African American, 2% American Indian/Alaska Native, 67% White and American Indian/Alaska Native
82	Ecuador	6.1% White, 5.3% Black/African American, 7% American Indian/Alaska Native, 1.9% White and Black/African American, 79.3% White and American Indian/Alaska Native, 0.4% Unrestricted Imputation QD05 when in combination with another race: 6.1% White, 5.3% Black/African American, 7.2% American Indian/Alaska Native, 1.9% White and Black/African American, 79.5% White and American Indian/Alaska Native
83	Peru	5.9% White, 3.6% Black/African American, 25.8% American Indian/Alaska Native, 0.6% Chinese, 0.6% Japanese, 60.2% White and American Indian/Alaska Native, 3.3% Unrestricted Imputation QD05 when in combination with another race: 5.9% White, 3.6% Black/African American, 27.5% American Indian/Alaska Native, 0.6% Chinese, 0.6% Japanese, 61.9% White and American Indian/Alaska Native

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
84	Bolivia	5% White, 1% Black/African American, 20% American Indian/Alaska Native, 70% White and American Indian/Alaska Native, 4% Unrestricted Imputation QD05 when in combination with another race: 5% White, 1% Black/African American, 22% American Indian/Alaska Native, 72% White and American Indian/Alaska Native
85	Chile	10.8% American Indian/Alaska Native, 88.9% White and American Indian/Alaska Native, 0.3% Unrestricted Imputation QD05 when in combination with another race: 10.8% American Indian/Alaska Native, 89.2% White and American Indian/Alaska Native
86	Argentina	94% White, 0.4% Black/African American, 2.4% American Indian/Alaska Native, 3.2% White and American Indian/Alaska Native
87	Paraguay	2.5% White, 2.5% American Indian/Alaska Native, 95% White and American Indian/Alaska Native
88	Uruguay	87.7% White, 4.6% Black/African American, 2.4% American Indian/Alaska Native, 5.3% Unrestricted Imputation QD05 when in combination with another race: 90.35% White, 7.25% Black/African American, 2.4% American Indian/Alaska Native
103	United Kingdom	87.2% White, 3% Black/African American, 4.2% Asian Indian, 2% Impute among Multiple Race, 3.6% Unrestricted Imputation
125	Saudi Arabia	QD05: 90% White, 10% Asian Indian
128	UAE	QD05: 21.8% White, 38.1% Asian Indian, 6.1% Filipino, 21.2% Other Asian, 12.8% Imputed among those who are not American Indian/Alaska Native
129	Qatar	QD05: 40% White, 36% Asian Indian, 10% Other Asian, 14% Imputed among those who are not American Indian/Alaska Native
130	Bahrain	QD05: 51.7% White, 1.6% Black/African American, 45.5% Asian Indian, 1.2% Impute among those who are not American Indian/Alaska Native
140	Kazakhstan	22.9% White, 72.7% Other Asian, 4.4% Imputed among those who are not American Indian/Alaska Native
141	Uzbekistan	6% White, 91% Other Asian, 3% Unrestricted Imputation
142	Tajikistan	2% White, 98% Other Asian
143	Kyrgyzstan	5.5% White, 89.3% Other Asian, 5.2% Imputed among those who are not American Indian/Alaska Native
144	Turkmenistan	4.0% White, 90.0% Other Asian, 6.0% Impute among those who are not American Indian/Alaska Native

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
157	Malaysia	6.9% Asian Indian, 23% Chinese, 69.2% Other Asian, 0.9% Imputed among those who are not American Indian/Alaska Native QD05 when in combination with another race: 6.9% Asian Indian, 23.45% Chinese, 69.65% Other Asian
159	Singapore	9% Asian Indian, 74.4% Chinese, 13.4% Other Asian, 3.2% Unrestricted Imputation
165	Djibouti	2.5% White, 97.5% Black/African American
166	Sudan	39% White, 58% Black/African American, 3% Impute among those who are not American Indian/Alaska Native
168	South Africa	7.8% White, 80.9% Black/African American, 1.3% Asian Indian, 1.3% Other Asian, 8.8% White and Black/African American
169	Namibia	6% White, 87.5% Black/African American, 6.5% White and Black/African American
170	Zimbabwe	99.4% Black/African American, 0.6% Imputed among those who are not American Indian/Alaska Native
171	Zambia	99% Black/African American, 1% Impute among those who are not American Indian/Alaska Native
173	Angola	1% White, 97% Black/African American, 2% White and Black/African American
175	Mauritius	QD05: 2% White, 68% Asian Indian, 3% Chinese, 27% White and Black/African American
177	Cape Verde	1% White, 28% Black/African American, 71% White and Black/African American
179	Mauritania	30% White, 30% Black/African American, 40% White and Black/African American
180	Mali	1.7% White, 92% Black/African American, 6.3% Imputed among those who are not American Indian/Alaska Native
181	Niger	11.4% White, 87.9% Black/African American, 0.7% Imputed among those who are not American Indian/Alaska Native
186	Australia	QD05: 74.3% White, 1.4% Asian Indian, 3.1% Chinese, 21.2% Imputed among those who are not American Indian/Alaska Native
187	New Zealand	QD05: 65% White, 3.9% Samoan, 20% Other Pacific Islander, 4.7% Asian Indian, 4.9% Chinese, 1.5% Filipino
902	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	13.5% White, 15.8% Black/African American, 41.4% White and Black/African American, 29% White and American Indian/Alaska Native, 0.3% Unrestricted Imputation
909	Creole	50% White, 50% White and Black/African American
915	Central/South American & West Indies	50% White and Black/African American, 50% Black/African American and American Indian/Alaska Native

Table C.3 Procedures for Codes Informative for Imputation

Race Code	Race Name	Restriction on Donors in Formal Imputation
33	Asian (nonspecific)	Donors were Asian: impute specific Asian group
36	Pacific Islander (nonspecific)	Donors were Native Hawaiian or Other Pacific Islander: impute specific Native Hawaiian or Other Pacific Islander group
54	Trinidad & Tobago (QD05: 15.3% Multiple Race)	Donors included respondents of more than one race
56	Virgin Is (St Thomas, St Croix) (QD05: 2.1% Multiple Race)	Donors included respondents of more than one race
70	Mexico	Donors were Mexican ¹
71	Puerto Rico	Donors were Puerto Rican
72	Cuba	Donors were Cuban
89	Mexico & Puerto Rico	Donors were Mexican, Puerto Rican, or both
90	Mexico & Cuba	Donors were Mexican, Cuban, or both
91	Mexico & Dominican Republic	Donors were Mexican, Dominican, or both
92	Mexico & Spain	Donors were Mexican, Spanish, or both
93	Puerto Rico & Cuba	Donors were Puerto Rican, Cuban, or both
94	Puerto Rico & Dominican Republic	Donors were Puerto Rican, Dominican, or both
95	Puerto Rico & Spain	Donors were Puerto Rican, Spanish, or both
96	Cuba & Dominican Republic	Donors were Cuban, Dominican, or both
97	Cuba & Spain	Donors were Cuban, Spanish, or both
98	Dominican Republic & Spain	Donors were Dominican, Spanish, or both
103	United Kingdom (QD05: 2% Multiple Race)	Donors included respondents of more than one race
128	UAE (QD05: 12.8% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
129	Qatar (QD05: 14% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
130	Bahrain (QD05: 1.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
140	Kazakhstan (QD05: 4.4% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
141	Uzbekistan (QD05: 3% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
143	Kyrgyzstan (QD05: 5.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
144	Turkmenistan (QD05: 6.0% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native

Table C.3 Procedures for Codes Informative for Imputation (continued)

Race Code	Race Name	Restriction on Donors in Formal Imputation
157	Malaysia (QD05: 0.09% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
159	Singapore (QD05: 3.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
166	Sudan (QD05: 3% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
170	Zimbabwe (QD05: 0.06% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
171	Zambia (QD05: 1% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
180	Mali (QD05: 6.3% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
181	Niger (QD05: 0.07% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
186	Australia (QD05: 21.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
201	Biracial (nonspecific)	Donors were multiple race: imputed constituent races ²
900	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain, D.R.)	Donors were Hispanic/Latino
901	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Donors were Hispanic/Latino
903	Central/South American (no country)	Donors were Central/South American
904	Nonwhite (nonspecific/brown)	Donors were any race but single-race White
905	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Donors were Hispanic/Latino who were any race but single-race White
906	Mezclado, Mezclada (Hispanic/Latino mixed)	Donors were multiple race and Hispanic/Latino: imputed constituent races
907	Mixed	Donors were multiple race: imputed constituent races
916	Central/South American & Mexican	Donors were Central/South American, Mexican, or both
917	Central/South American & Puerto Rican	Donors were Central/South American, Puerto Rican, or both
918	Central/South American & Cuban	Donors were Central/South American, Cuban, or both
919	Central/South American & Dominican	Donors were Central/South American, Dominican, or both
920	Central/South American & Spanish	Donors were Central/South American, Spanish, or both

Table C.3 Procedures for Codes Informative for Imputation (continued)

Race Code	Race Name	Restriction on Donors in Formal Imputation
932	Puerto Rican & Dominican	Donors were Puerto Rican, Dominican, or both
933	Puerto Rican & Spanish (from Spain)	Donors were Puerto Rican, Spanish, or both
935	Cuban & Spanish (from Spain)	Donors were Cuban, Spanish, or both
936	Dominican & Spanish (from Spain)	Donors were Dominican, Spanish, or both

¹ Even though recipients may not be Hispanic/Latino, they may still have indicated “Mexican” in the QD05 “OTHER, Specify” response. Donors in this instance included both Hispanic/Latino and (though extremely rare) non-Hispanic/Latino Mexicans.

² Because most multiple-race respondents have only two constituent races, any respondent with this code and nothing else is likely to be assigned a biracial donor. However, for the sake of simplicity, respondents with this code were not treated any differently than respondents with code 907 (“Mixed”).

Table C.4 Mapping of Hispanic/Latino Group Codes

Hispanic/ Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/ Latino Code Directly Mapped
11	Mexican/Mexican American/Mexicano/Chicano	Mexican
12	Puerto Rican	Puerto Rican
13	Central or South American	Central or South American
14	Cuban/Cuban American	Cuban
15	Dominican (Dominican Republic)	Dominican
16	Spanish (from Spain)	Spanish
17	Caribbean Hispanic/Latino (not specified as Dominican)	Other Hispanic
21	Mexican & Puerto Rican	Mexican, Puerto Rican
22	Mexican & Central or South American	Mexican, Central or South American
23	Mexican & Cuban	Mexican, Cuban
24	Mexican & Dominican	Mexican, Dominican
25	Mexican & Spanish (from Spain)	Mexican, Spanish
26	Puerto Rican & Central or South American	Puerto Rican, Central or South American
27	Puerto Rican & Cuban	Puerto Rican, Cuban
28	Puerto Rican & Dominican	Puerto Rican, Dominican
29	Puerto Rican & Spanish (from Spain)	Puerto Rican, Spanish
30	Central or South American & Cuban	Central or South American, Cuban
31	Central or South American & Dominican	Central or South American, Dominican
32	Central or South American & Spanish (from Spain)	Central or South American, Spanish
33	Cuban & Dominican	Cuban, Dominican
34	Cuban & Spanish (from Spain)	Cuban, Spanish
35	Dominican & Spanish (from Spain)	Dominican, Spanish
36	Mexican, Puerto Rican, & Central or South American	Mexican, Puerto Rican, Central or South American
37	Mexican, Puerto Rican, & Cuban	Mexican, Puerto Rican, Cuban
38	Mexican, Puerto Rican, & Dominican	Mexican, Puerto Rican, Dominican
39	Mexican, Puerto Rican, & Spanish (from Spain)	Mexican, Puerto Rican, Spanish
40	Mexican, Central or South American, & Cuban	Mexican, Central or South American, Cuban
41	Mexican, Central or South American, & Dominican	Mexican, Central or South American, Dominican
42	Mexican, Central or South American, & Spanish (from Spain)	Mexican, Central or South American, Spanish

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/ Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/ Latino Code Directly Mapped
43	Mexican, Cuban, & Dominican	Mexican, Cuban, Dominican
44	Mexican, Cuban, & Spanish (from Spain)	Mexican, Cuban, Spanish
45	Mexican, Dominican, & Spanish (from Spain)	Mexican, Dominican, Spanish
46	Puerto Rican, Central or South American, & Cuban	Puerto Rican, Central or South American, Cuban
47	Puerto Rican, Central or South American, & Dominican	Puerto Rican, Central or South American, Dominican
48	Puerto Rican, Central or South American, & Spanish (from Spain)	Puerto Rican, Central or South American, Spanish
49	Puerto Rican, Cuban, & Dominican	Puerto Rican, Cuban, Dominican
50	Puerto Rican, Cuban, & Spanish (from Spain)	Puerto Rican, Cuban, Spanish
51	Puerto Rican, Dominican, & Spanish (from Spain)	Puerto Rican, Dominican, Spanish
52	Central or South American, Cuban, & Dominican	Central or South American, Cuban, Dominican
53	Central or South American, Cuban, & Spanish (from Spain)	Central or South American, Cuban, Spanish
54	Central or South American, Dominican, & Spanish (from Spain)	Central or South American, Dominican, Spanish
55	Cuban, Dominican, and Spanish (from Spain)	Cuban, Dominican, Spanish
56	Portuguese & Mexican	Mexican, Other Hispanic
57	Portuguese & Puerto Rican	Puerto Rican, Other Hispanic
58	Portuguese & Cuban	Cuban, Other Hispanic
59	Portuguese & Central or South American	Central or South American, Other Hispanic
60	Portuguese & Dominican	Dominican, Other Hispanic
61	Portuguese & Spanish (from Spain)	Spanish, Other Hispanic
62	Spain & Philippines	Spanish, Other Hispanic
100	Brazilian	Central or South American
101	Portuguese	Other Hispanic/Latino
102	Cape Verde	Other Hispanic/Latino
103	Belizean (formerly British Honduras)	Central or South American
104	Guyana	Central or South American
105	Jamaican	Other Hispanic/Latino
106	Other Caribbean (possibly Hispanic)	Other Hispanic/Latino

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/ Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/ Latino Code Directly Mapped
107	Philippines & Guam	Other Hispanic/Latino
108	Brazilian & Portuguese	Central or South American
109	Cape Verde & Portuguese	Other Hispanic/Latino
110	Haitian/Jamaican	Other Hispanic/Latino
200	Mexican/Jamaican	Mexican, Other Hispanic/Latino
201	Puerto Rican/Jamaican	Puerto Rican, Other Hispanic/Latino
202	Central or South American/Jamaican	Central or South American, Other Hispanic/Latino
203	Cuban/Jamaican	Cuban, Other Hispanic/Latino
204	Dominican/Jamaican	Dominican, Other Hispanic/Latino
205	Spanish (from Spain)/Jamaican	Spanish, Other Hispanic/Latino
206	Mexican/West Indies	Mexican, Other Hispanic/Latino
207	Puerto Rican/West Indies	Puerto Rican, Other Hispanic/Latino
208	Central or South American/West Indies	Central or South American, Other Hispanic/Latino
209	Cuban/West Indies	Cuban, Other Hispanic/Latino
210	Dominican/West Indies	Dominican, Other Hispanic/Latino
211	Spanish (from Spain)/West Indies	Spanish, Other Hispanic/Latino
212	Mexican/Haitian	Mexican, Other Hispanic/Latino
213	Puerto Rican/Haitian	Puerto Rican, Other Hispanic/Latino
214	Central or South American/Haitian	Central or South American, Other Hispanic/Latino
215	Cuban/Haitian	Cuban, Other Hispanic/Latino
216	Dominican/Haitian	Dominican, Other Hispanic/Latino
217	Spanish (from Spain)/Haitian	Spanish, Other Hispanic/Latino
500	Hispanic/Latino	Hispanic/Latino group imputed
501	Hispanic/Latino Mixed/Mezclada	Hispanic group imputed
502	Hispanic Creole	Other Hispanic
600*	Stated clearly as Not Hispanic/Latino	Hispanic/Latino indicator edited to "no"
800	Non-Hispanic/Latino country	Other Hispanic/Latino
801	Racial category (White, Black/African American, etc.)	Hispanic group imputed
802	Combination race and non-Hispanic country	Other Hispanic/Latino
985	Bad Data/"Mixed"	Hispanic/Latino group imputed
994	Unknown/"Don't Know"	Hispanic/Latino group imputed
997	American or "All of Them"	Hispanic/Latino group imputed

*This code caused the Hispanic/Latino indicator to be edited to a "no." Codes that caused the Hispanic/Latino indicator to be edited to a "yes" are listed in [Table C.1](#).

Appendix D: Model Summaries

D.1 Introduction

The following exhibits in this appendix present the lists of available covariates for the imputation models that were run in the 2023 National Survey on Drug Use and Health (NSDUH):

- [Exhibit D.1](#): Demographic Variables,
- [Exhibit D.2](#): Substance Use Variables,
- [Exhibit D.3](#): Mental Health Variables,
- [Exhibit D.4](#): Substance Use Treatment Variables,
- [Exhibit D.5](#): Mental Health Treatment Variables,
- [Exhibit D.6](#): Household Composition Variables,
- [Exhibit D.7](#): Income Variables,
- [Exhibit D.8](#): Health Insurance Variables, and
- [Exhibit D.9](#): Pair Relationship Variables.

Note that because imputation is sequential, not all covariates listed in a particular exhibit are available for each outcome being imputed. For example, the first demographic variable to undergo imputation is marital status. Therefore, variables that are imputed after marital status (e.g., race, employment status) are not available to use as covariates in the marital status models. More information about the sequence of imputation can be found in Chapters 4 through 14.

In the exhibits, the variable “age” is the mean-centered age, where the age was “centered” by subtracting the mean age and where the mean was calculated across all respondents within the age group who were used to build the given model. The variables “age squared” and “age cubed” represent the square and cube, respectively, of this mean-centered age variable. The variable “Orthogonal Age” is the orthogonal polynomial term for age calculated across all respondents within the domain. The variable “Orthogonal Age Squared” is also an orthogonal polynomial term. Using orthogonal polynomials reduces the severity of multicollinearity in the regression models. When an asterisk “*” is included in the exhibits, it represents an interaction between two variables.

D.2 Screener, Block-Level, and Segment-Level Variables

In the predictive mean neighborhood (PMN) and modified predictive mean neighborhood (modPMN) procedures, statistical modeling was performed to adjust weights for item nonresponse and to calculate predicted means for various outcomes. Descriptions of questionnaire-derived variables are described in the main body of this report, with the exception of screener, block-level, and segment-level variables, which are described below. These variables were often used as covariates in both types of models for the PMN and modPMN procedures.

D.2.1 Household Type

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

D.2.2 Census Region

Census region was a four-level geographic variable recoded from the respondent's state of residence. The four levels were Northeast, Midwest, South, and West. The Northeast includes Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Midwest consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia compose the South. Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming compose the West.

D.2.3 Population Density

The population density variable classifies respondents according to their living situation, whether it be in a rural or urban area and, if urban, the size of the urban area. It was used to categorize segments where the respondents lived according to modified 2010 census data, which were adjusted to reflect population increases between census years by intercensal projections by Claritas, Inc.¹ This variable had five levels: segment in core-based statistical area (CBSA)² with 1 million or more people; segment in CBSA with 250,000 to 999,999 people; segment in CBSA with fewer than 250,000 people; segment not in CBSA and not in rural area; and segment not in CBSA and in rural area.

D.2.4 Percentage Hispanic/Latino in Block

The "Percentage Hispanic/Latino in Block" variable was used to categorize blocks according to the concentration of Hispanics/Latinos in the blocks in which the respondents lived, using the adjusted 2010 census data. It had three levels: less than 20 percent, 20 to less than 71 percent, and 71 percent or more.

D.2.5 Percentage Owner Occupied in Block

The "Percentage Owner Occupied in Block" variable was used to categorize blocks according to the concentration of owner-occupied households in the blocks in which the respondents lived, using the adjusted 2010 census data. It was used as a surrogate for income because wealthy

¹ Claritas, Inc. is a market research firm headquartered in San Diego, California.

² [CBSAs](#), developed in response to standards put forth by the Office of Management and Budget (OMB), are metropolitan and micropolitan areas that were designated using data from the 2000 census.

blocks tend to have many homeowners, whereas poor blocks tend to have many renters. It had three levels: less than 10 percent, 10 to less than 50 percent, and 50 percent or more.

D.2.6 Percentage Black/African American in Block

The “Percentage Black/African American in Block” variable was used to categorize blocks according to the concentration of Black/African American households in the blocks in which the respondents lived, using the adjusted 2010 census data. It had three levels: less than 10 percent, 10 to less than 40 percent, and 40 percent or more.

D.2.7 Percentage Asian/Other Pacific Islander in Block

The “Percentage Asian/Other Pacific Islander in Block” variable was used to categorize blocks according to the concentration of Asian/Other Pacific Islander households in the blocks in which the respondents lived, using the adjusted 2010 census data. It had three levels: less than 5 percent, 5 to less than 10 percent, and 10 percent or more.

D.2.8 Percentage American Indian/Alaska Native in Block

The “Percentage American Indian/Alaska Native in Block” variable was used to categorize blocks according to the concentration of American Indian/Alaska Native households in the blocks in which the respondents lived, using the adjusted 2010 census data. It had three levels: less than 1 percent, 1 to less than 3 percent, and 3 percent or more.

D.2.9 Percentage Non-Hispanic/Latino Black/African American in Block

The “Percentage Non-Hispanic/Latino Black/African American in Block” variable was used to categorize blocks according to the concentration of non-Hispanic/Latino Black/African American households in the blocks in which the respondents lived, using the adjusted 2010 census data. It had three levels: less than 10 percent, 10 percent to less than 50 percent, and 50 percent or more.

D.3 Definitions of Levels for Categorical Variables

Exhibit D.1 Demographic Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category

12-17, 18-25, 26-34, 35-49, 50+¹

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Race/Hispanic Recode (6 Levels)

White Only,¹ Black/African American Only, American Indian/Alaska Native Only, Hawaiian/Pacific Islander Only, Asian, Multiple Race

Exhibit D.1 Demographic Variables (continued)

Race/Hispanic Recode (9 Levels)

Not Hispanic/Latino and White,¹ Not Hispanic/Latino and Black/African American, Not Hispanic/Latino and American Indian/Alaska Native, Not Hispanic/Latino and Hawaiian/Pacific Islander, Not Hispanic/Latino and Asian, Not Hispanic/Latino and Multiple Race, Hispanic/Latino and Puerto Rican, Hispanic/Latino and Mexican, Hispanic/Latino Other

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Immigrant Education Level (4 Levels)

Less Than High School, Includes 12- to 17-Year-Olds, High School Graduate, Some College, College Graduate¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Immigrant Employment Status

Full Time, Part Time, Unemployed, Other Employment or Aged 12-17¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Immigrant Marital Status (3 Levels)

Married, Was Married, Never Been Married or Aged 12-17¹

Marital Status (4 Levels)

Married, Widowed, Divorced or Separated, Never Been Married¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Population Density (5 Levels)

Segment in CBSA with 1 Million or More People, Segment in CBSA from 250,000 to 999,999 People, Segment in CBSA with Fewer Than 250,000 People, Segment Not in CBSA and Not in a Rural Area, Segment Not in CBSA and in a Rural Area¹

Census Region

Northeast, Midwest, South, West¹

Race/Ethnicity of Householder

Black/African American, Hispanic/Latino, White¹

Percentage Asian/Other Pacific Islander in Block

≥ 10%, [5,10)%, < 5%¹

Percentage Black/African American in Block

≥ 40%, [10,40)%, < 10%¹

Percentage Hispanic/Latino in Block

≥ 71%, [20,71)%, < 20%¹

Percentage American Indian/Alaska Native in Block

≥ 3%, [1,3)%, < 1%¹

Percentage Owner Occupied in Block

≥ 50%, [10,50)%, < 10%¹

Age Category * Gender

Male 12-17, Male 18-25, Male 26-34, Male 35-49, Male 50+,¹ Female 12-17,¹ Female 18-25,¹ Female 26-34,¹ Female 35-49,¹ Female 50+¹

Exhibit D.1 Demographic Variables (continued)

Gender * Race/Hispanic Recode (9 Levels)

Male Not Hispanic/Latino and White,¹ Male Not Hispanic/Latino and Black/African American, Male Not Hispanic/Latino and American Indian/Alaska Native, Male Not Hispanic/Latino and Hawaiian/Pacific Islander, Male Not Hispanic/Latino and Asian/Other Pacific Islander, Male Not Hispanic/Latino and Multiple Race, Male Hispanic/Latino and Puerto Rican, Male Hispanic/Latino and Mexican, Male Hispanic/Latino Other, Female Not Hispanic/Latino and White,¹ Female Not Hispanic/Latino and Black/African American,¹ Female Not Hispanic/Latino and American Indian/Alaska Native,¹ Female Not Hispanic/Latino and Hawaiian/Pacific Islander,¹ Female Not Hispanic/Latino and Asian/Other Pacific Islander,¹ Female Not Hispanic/Latino and Multiple Race,¹ Female Hispanic/Latino and Puerto Rican,¹ Female Hispanic/Latino and Mexican,¹ Female Hispanic/Latino Other¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.2 Substance Use Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category (2 Levels)

12-17,¹ 18+

Age Category - Adults (2 Levels)

26-49, 50+¹

Age Category (3 Levels)

12-17,¹ 18-25, 26+

Age Category - Adults (3 Levels)

26-34, 35-49, 50+¹

Age Category (4 Levels)

12-17,¹ 18-25, 26-49, 50+

Age Category (5 Levels)

12-17,¹ 18-25, 26-34, 35-49, 50+

Race/Hispanic Recode (2 Levels)

Not Hispanic/Latino and White Only,¹ Hispanic/Latino or Not White Only

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Education Level (2 Levels)

Not a College Graduate, College Graduate¹

Education Level (3 Levels)

Less Than High School, High School Graduate/Some College, College Graduate¹

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (2 Levels)

Full/Part Time, Unemployed/Other Employment¹

Employment Status (2 Levels) - Heroin 30-Day Frequency Models

Full Time, Part Time/Unemployed/Other Employment¹

Employment Status (3 Levels)

Full Time, Part Time, Unemployed/Other Employment¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Exhibit D.2 Substance Use Variables (continued)

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Population Density (2 Levels)

Segment Not in CBSA or Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Population Density (2 Levels) - Heroin 30-Day Frequency Models

Segment Not in CBSA, Segment in CBSA¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Census Region

Northeast, Midwest, South, West¹

Gender * Race/Hispanic Recode (2 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Other, Female Not Hispanic/Latino and White Only,¹ Female Other¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Age * Race/Hispanic Recode (2 Levels)

Age * Not Hispanic/Latino and White Only,¹ Age * Hispanic/Latino or Not White Only

Age * Race/Hispanic Recode (4 Levels)

Age * Not Hispanic/Latino and White Only,¹ Age * Not Hispanic/Latino and Black/African American Only, Age * Not Hispanic/Latino and Other, Age * Hispanic/Latino

Age Squared * Race/Hispanic Recode (2 Levels)

Age Squared * Not Hispanic/Latino and White Only,¹ Age Squared * Hispanic/Latino or Not White Only

Age Squared * Race/Hispanic Recode (4 Levels)

Age Squared * Not Hispanic/Latino and White Only,¹ Age Squared * Not Hispanic/Latino and Black/African American Only, Age Squared * Not Hispanic/Latino and Other, Age Squared * Hispanic/Latino

Orthogonal Age * Race/Hispanic Recode (2 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Hispanic/Latino or Not White Only

Orthogonal Age * Race/Hispanic Recode (4 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Not Hispanic/Latino and Black/African American Only, Orthogonal Age * Not Hispanic/Latino and Other, Orthogonal Age * Hispanic/Latino

Orthogonal Age Squared * Race/Hispanic Recode (2 Levels)

Orthogonal Age Squared * Not Hispanic/Latino and White Only,¹ Orthogonal Age Squared * Hispanic/Latino or Not White Only

Orthogonal Age Squared * Race/Hispanic Recode (4 Levels)

Orthogonal Age Squared * Not Hispanic/Latino and White Only,¹ Orthogonal Age Squared * Not Hispanic/Latino and Black/African American Only, Orthogonal Age Squared * Not Hispanic/Latino and Other, Orthogonal Age Squared * Hispanic/Latino

Cigarette Lifetime Indicator

Yes, No¹

Intermediate Drug-Specific Lifetime Use Indicator

Yes, No¹

Drug-Specific Lifetime Use Indicator

Yes, No¹

Exhibit D.2 Substance Use Variables (continued)

Intermediate Marijuana Past Year Way of Use Indicator

Yes, No¹

Drug-Specific Past Year Misuse Indicator

Yes, No¹

Drug-Specific Past Year Use Indicator

Yes, No¹

Intermediate Drug-Specific Past Month Use Indicator

Yes, No¹

Intermediate Marijuana Past Month Way of Use Indicator

Yes, No¹

Drug-Specific Past Month Use Indicator

Yes, No¹

Pipe Recency

Past Month, Lifetime but Not Past Month, Lifetime Nonuser¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Past Year, Lifetime but Not Past Year, Lifetime Nonuser¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)

Past Month, Past Year but Not Past Month, Lifetime but Not Past Year, Lifetime Nonuser¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)

Past Month, Past Year but Not Past Month, Past 3 Years but Not Past Year, Lifetime but Not Past 3 Years, Lifetime Nonuser¹

Drug-Specific Recency, Domain Restricted to Lifetime Users (2 Levels)

Past Year, Lifetime but Not Past Year¹

Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels)

Past Month, Past Year but Not Past Month, Lifetime but Not Past Year¹

Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels)

Past Month, Past Year but Not Past Month, Past 3 Years but Not Past Year, Lifetime but Not Past 3 Years¹

Drug-Specific Any Use Recency, Domain Not Restricted to Lifetime Users (1 Level)

Lifetime Any Use, Never Used¹

Drug-Specific Any Use Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Past Year Any Use, Lifetime but Not Past Year Any Use, Never Used¹

Past Year Any Use of Sedatives or Tranquilizers

Sedatives Only, Tranquilizers Only, Sedatives and Tranquilizers¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Lifetime Misuse, Never Misused¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (3 Levels)

Lifetime Misuse, Lifetime Any Use but Not Misuse, Never Used¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (4 Levels)

Past Month Misuse, Past Year but Not Past Month Misuse, Lifetime but Not Past Year Misuse, Lifetime Any Use but Not Misuse, Never Used¹

Drug-Specific Past Year Initiation

Yes, No¹

Drug-Specific Past Year Use Disorder Indicator

Yes, No¹

Drug-Specific Past Year Misuse Disorder Indicator

Yes, No¹

Drug-Specific Past Year Use but Not Misuse Disorder Indicator

Yes, No¹

Drug-Specific Past Year Use Disorder Severity (2 Levels)

Mild Disorder, Moderate or Severe Disorder, No Disorder¹

Drug-Specific Past Year Misuse Disorder Severity (2 Levels)

Mild Disorder, Moderate or Severe Disorder, No Disorder¹

Exhibit D.2 Substance Use Variables (continued)

Drug-Specific Past Year Use but Not Misuse Disorder Severity (2 Levels)

Mild Disorder, Moderate or Severe Disorder, No Disorder¹

Drug-Specific Disorder Severity (3 Levels)

Mild Disorder, Moderate Disorder, Severe Disorder, No Disorder¹

Drug-Specific Past Year Misuse Disorder Severity (3 Levels)

Mild Disorder, Moderate Disorder, Severe Disorder, No Disorder¹

Drug-Specific Past Year Use but Not Misuse Disorder Severity (3 Levels)

Mild Disorder, Moderate Disorder, Severe Disorder, No Disorder¹

Lifetime State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Lifetime State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Hallucinogen Lifetime State Rank (2 Levels)²

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Hallucinogen Lifetime State Rank (3 Levels)²

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Any Use Recency State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Any Use Recency State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Recency State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Recency State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Initiation State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Initiation State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

² For lifetime hallucinogen child drug models, hallucinogen lifetime state rank was used rather than the lifetime state rank of the individual child drugs due to the small number of lifetime users for some child drugs.

Exhibit D.3 Mental Health Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category - Adults (3 Levels)

18-25, 26-49, 50+¹

Age Category - Adults (4 Levels)

18-25, 26-34, 35-49, 50+¹

Age Category - Adults (5 Levels)

18-25, 26-34, 35-49, 50-64, 65+¹

Race/Hispanic Recode (2 Levels)

Not Hispanic/Latino and White Only,¹ Hispanic/Latino or Not White Only

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Exhibit D.3 Mental Health Variables (continued)

Education Level (2 Levels)

Not a College Graduate, College Graduate¹

Education Level (3 Levels)

Less Than High School, High School Graduate/Some College, College Graduate¹

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (2 Levels)

Full/Part Time, Unemployed/Other Employment¹

Employment Status (3 Levels)

Full Time, Part Time, Unemployed/Other Employment¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Population Density (2 Levels)

Segment Not in CBSA or Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Census Region

Northeast, Midwest, South, West¹

Gender * Race/Hispanic Recode (2 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Other, Female Not Hispanic/Latino and White Only,¹ Female Other¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Orthogonal Age * Race/Hispanic Recode (2 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Hispanic/Latino or Not White Only

Orthogonal Age * Race/Hispanic Recode (4 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Not Hispanic/Latino and Black/African American Only, Orthogonal Age * Not Hispanic/Latino and Other, Orthogonal Age * Hispanic/Latino

Cigarette Lifetime Indicator

Yes, No¹

Drug-Specific Lifetime Use Indicator

Yes, No¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Past Year, Lifetime but Not Past Year, Lifetime Nonuser¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)

Past Month, Past Year but Not Past Month, Lifetime but Not Past Year, Lifetime Nonuser¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)

Past Month, Past Year but Not Past Month, Past 3 Years but Not Past Year, Lifetime but Not Past 3 Years, Lifetime Nonuser¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Lifetime Misuse, Never Misused¹

Exhibit D.3 Mental Health Variables (continued)

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (3 Levels)

Lifetime Misuse, Lifetime Any Use but Not Misuse, Never Used¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (4 Levels)

Past Month Misuse, Past Year but Not Past Month Misuse, Lifetime but Not Past Year Misuse, Lifetime Any Use but Not Misuse, Never Used¹

Drug-Specific Past Year Use Disorder Indicator

Yes, No¹

Drug-Specific Past Year Any Use Disorder Indicator

Yes, No¹

Participated in One or More Government Assist Programs

Yes, No¹

Total Family Income (3 Levels)

< \$50,000, \$50,000 - \$99,999, \$100,000 or more¹

Total Family Income (5 Levels)

< \$20,000, \$20,000 - \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 or more¹

Medicare Coverage

Yes, No¹

Medicaid/CHIP Coverage

Yes, No¹

CHAMPUS Coverage

Yes, No¹

Private Health Insurance Coverage

Yes, No¹

Other Health Insurance Coverage

Yes, No¹

Month in Past Year Felt Worse Than Past 30 Days

Yes, No¹

Past Year Serious Psychological Distress Indicator

Yes, No¹

Serious Thoughts of Suicide

Yes, No¹

Made Suicide Plans

Yes, No¹

Attempted Suicide

Yes, No¹

Serious Thoughts of Suicide due to COVID-19

Yes, No¹

Made Suicide Plans due to COVID-19

Yes, No¹

Attempted Suicide due to COVID-19

Yes, No¹

Kessler-6 State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Kessler-6 State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

WHODAS State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

WHODAS State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Major Depressive Episode State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Major Depressive Episode State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Exhibit D.3 Mental Health Variables (continued)

Suicide Behavior State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Suicide Behavior State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Suicide Behavior due to COVID-19 State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Suicide Behavior due to COVID-19 State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Medical Attention due to Suicide Attempt State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Medical Attention due to Suicide Attempt State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

Hospital Stay due to Suicide Attempt State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Hospital Stay due to Suicide Attempt State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area; COVID-19 = coronavirus disease 2019.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.4 Substance Use Treatment Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category (2 Levels)

12-17,¹ 18+

Age Category (3 Levels)

12-17,¹ 18-25, 26+

Age Category (4 Levels)

12-17,¹ 18-25, 26-49, 50+

Age Category (5 Levels)

12-17,¹ 18-25, 26-34, 35-49, 50+

Race/Hispanic Recode (2 Levels)

Not Hispanic/Latino and White Only,¹ Hispanic/Latino or Not White Only

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Education Level (2 Levels)

Not a College Graduate, College Graduate¹

Education Level (3 Levels)

Less Than High School, High School Graduate/Some College, College Graduate¹

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (2 Levels)

Full/Part Time, Unemployed/Other Employment¹

Employment Status (3 Levels)

Full Time, Part Time, Unemployed/Other Employment¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Exhibit D.4 Substance Use Treatment Variables (continued)

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Gender * Race/Hispanic Recode (2 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Other, Female Not Hispanic/Latino and White Only,¹ Female Other¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Orthogonal Age * Race/Hispanic Recode (2 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Hispanic/Latino or Not White Only

Orthogonal Age * Race/Hispanic Recode (4 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Not Hispanic/Latino and Black/African American Only, Orthogonal Age * Not Hispanic/Latino and Other, Orthogonal Age * Hispanic/Latino

Drug-Specific Lifetime Use Indicator

Yes, No¹

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Past Year, Lifetime but Not Past Year, Lifetime Nonuser¹

Drug-Specific Recency, Domain Restricted to Lifetime Users (2 Levels)

Past Year, Lifetime but Not Past Year¹

Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels)

Past Month, Past Year but Not Past Month, Lifetime but Not Past Year¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (2 Levels)

Lifetime Misuse, Never Misused¹

Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (3 Levels)

Lifetime Misuse, Lifetime Any Use but Not Misuse, Never Used¹

Drug-Specific Past Year Use Disorder Indicator

Yes, No¹

Drug-Specific Past Year Any Use Disorder Indicator

Yes, No¹

Drug-Specific Past Year Use Disorder Severity (2 Levels)

Mild Disorder, Moderate or Severe Disorder, No Disorder¹

Drug-Specific Past Year Any Use Disorder Severity (2 Levels)

Mild Disorder, Moderate or Severe Disorder, No Disorder¹

Participated in One or More Government Assist Programs

Yes, No¹

Total Family Income (3 Levels)

< \$50,000, \$50,000 - \$99,999, \$100,000 or more¹

Total Family Income (5 Levels)

< \$20,000, \$20,000 - \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 or more¹

Medicare Coverage

Yes, No¹

CHAMPUS Coverage

Yes, No¹

Medicaid/CHIP Coverage

Yes, No¹

Exhibit D.4 Substance Use Treatment Variables (continued)

Private Health Insurance Coverage

Yes, No¹

Other Health Insurance Coverage

Yes, No¹

Location/Type of SUT Received: Hospital as an Inpatient

Yes, No¹

Location/Type of SUT Received: Inpatient Rehab

Yes, No¹

Location/Type of SUT Received: Residential Mental Health Treatment Center

Yes, No¹

Location/Type of SUT Received: Other Place as Inpatient

Yes, No¹

Location/Type of SUT Received: General Medical Office

Yes, No¹

Location/Type of SUT Received: Hospital as an Outpatient

Yes, No¹

Location/Type of SUT Received: Outpatient Mental Health Treatment Center

Yes, No¹

Location/Type of SUT Received: Outpatient Rehab

Yes, No¹

Location/Type of SUT Received: School Health/Counseling

Yes, No¹

Location/Type of SUT Received: Other Place as Outpatient

Yes, No¹

Location/Type of SUT Received: Virtual Services

Yes, No¹

Location/Type of SUT Received: Prison

Yes, No¹

Used Prescription to Help Cut Back or Stop Alcohol Use

Yes, No¹

Used Prescription to Help Cut Back or Stop Heroin/Pain Reliever Use

Yes, No¹

Location/Type of SUT Received: Therapist Office

Yes, No¹

Location/Type of SUT Received: Emergency Room

Yes, No¹

Location/Type of SUT Received: Support Group

Yes, No¹

Location/Type of SUT Received: Peer Support Specialist

Yes, No¹

Location/Type of SUT Received: Detox Services

Yes, No¹

SUT State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

SUT State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area; SUT = substance use treatment.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.5 Mental Health Treatment Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category (2 Levels)

12-17,¹ 18+

Age Category (3 Levels)

12-17,¹ 18-25, 26+

Age Category (4 Levels)

12-17,¹ 18-25, 26-49, 50+

Age Category (5 Levels)

12-17,¹ 18-25, 26-34, 35-49, 50+

Race/Hispanic Recode (2 Levels)

Not Hispanic/Latino and White Only,¹ Hispanic/Latino or Not White Only

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Education Level (2 Levels)

Not a College Graduate, College Graduate¹

Education Level (3 Levels)

Less Than High School, High School Graduate/Some College, College Graduate¹

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (2 Levels)

Full/Part Time, Unemployed/Other Employment¹

Employment Status (3 Levels)

Full Time, Part Time, Unemployed/Other Employment¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Gender * Race/Hispanic Recode (2 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Other, Female Not Hispanic/Latino and White Only,¹ Female Other¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Orthogonal Age * Race/Hispanic Recode (2 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Hispanic/Latino or Not White Only

Orthogonal Age * Race/Hispanic Recode (4 Levels)

Orthogonal Age * Not Hispanic/Latino and White Only,¹ Orthogonal Age * Not Hispanic/Latino and Black/African American Only, Orthogonal Age * Not Hispanic/Latino and Other, Orthogonal Age * Hispanic/Latino

Drug-Specific Lifetime Use Indicator

Yes, No¹

Exhibit D.5 Mental Health Treatment Variables (continued)

Drug-Specific Recency, Domain Not Restricted to Lifetime Users (2 Levels)
Past Year, Lifetime but Not Past Year, Lifetime Nonuser ¹
Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (2 Levels)
Lifetime Misuse, Never Misused ¹
Drug-Specific Misuse Recency, Domain Not Restricted to Lifetime Users (3 Levels)
Lifetime Misuse, Lifetime Any Use but Not Misuse, Never Used ¹
Drug-Specific Past Year Use Disorder Indicator
Yes, No ¹
Drug-Specific Past Year Any Use Disorder Indicator
Yes, No ¹
Drug-Specific Past Year Use Disorder Severity (2 Levels)
Mild Disorder, Moderate or Severe Disorder, No Disorder ¹
Drug-Specific Past Year Any Use Disorder Severity (2 Levels)
Mild Disorder, Moderate or Severe Disorder, No Disorder ¹
Participated in One or More Government Assist Programs
Yes, No ¹
Total Family Income (3 Levels)
< \$50,000, \$50,000 - \$99,999, \$100,000 or more ¹
Total Family Income (5 Levels)
< \$20,000, \$20,000 - \$49,999, \$50,000 - \$74,999, \$75,000 - \$99,999, \$100,000 or more ¹
Medicare Coverage
Yes, No ¹
CHAMPUS Coverage
Yes, No ¹
Medicaid CHIP Coverage
Yes, No ¹
Private Health Insurance Coverage
Yes, No ¹
Other Health Insurance Coverage
Yes, No ¹
Adult Serious Psychological Distress
Yes, No or N/A ¹
Adult Lifetime Major Depressive Episode
Yes, No or N/A ¹
Adult Past Year Major Depressive Episode
Yes, No or N/A ¹
Adult Major Depressive Episode Impairment
Yes, No or N/A ¹
Adult Serious Thoughts of Suicide
Yes, No or N/A ¹
Adult Made Suicide Plans
Yes, No or N/A ¹
Adult Attempted Suicide
Yes, No or N/A ¹
Youth Lifetime Major Depressive Episode
Yes, No or Unknown or N/A ¹
Youth Past Year Major Depressive Episode
Yes, No or Unknown or N/A ¹
Youth Serious Thoughts of Suicide
Yes, Not sure/don't know, Don't want to answer/refused, No or Unknown or N/A ¹
Youth Made Suicide Plans
Yes, Not sure/don't know, Don't want to answer/refused, No or Unknown or N/A ¹

Exhibit D.5 Mental Health Treatment Variables (continued)

Youth Attempted Suicide

Yes, Not sure/don't know, Don't want to answer/refused, No or Unknown or N/A¹

Location/Type of MH Treatment Received: Hospital as an Inpatient

Yes, No¹

Location/Type of MH Treatment Received: Inpatient Rehab

Yes, No¹

Location/Type of MH Treatment Received: Residential Mental Health Treatment Center

Yes, No¹

Location/Type of MH Treatment Received: Other Place as Inpatient

Yes, No¹

Location/Type of MH Treatment Received: General Medical Office

Yes, No¹

Location/Type of MH Treatment Received: Hospital as an Outpatient

Yes, No¹

Location/Type of MH Treatment Received: Outpatient Mental Health Treatment Center

Yes, No¹

Location/Type of MH Treatment Received: Outpatient Rehab

Yes, No¹

Location/Type of MH Treatment Received: School Health/Counseling

Yes, No¹

Location/Type of MH Treatment Received: Other Place as Outpatient

Yes, No¹

Location/Type of MH Treatment Received: Virtual Services

Yes, No¹

Location/Type of MH Treatment Received: Prison

Yes, No¹

Location/Type of MH Treatment Received: Prescription Medication

Yes, No¹

Location/Type of MH Treatment Received: Therapist Office

Yes, No¹

Location/Type of MH Treatment Received: Emergency Room

Yes, No¹

Location/Type of MH Treatment Received: Support Group

Yes, No¹

Location/Type of MH Treatment Received: Peer Support Specialist

Yes, No¹

Mental Health Treatment State Rank (2 Levels)

Low/Middle State Rank (low/middle tertile), High State Rank (high tertile)¹

Mental Health Treatment State Rank (3 Levels)

Low State Rank (low tertile), Middle State Rank (middle tertile), High State Rank (high tertile)¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area; MH = mental health.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.6 Household Composition Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Age Category (5 Levels)²

12-17,¹ 18-25,¹ 26-34, 35-49, 50+¹

Exhibit D.6 Household Composition Variables (continued)

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Census Region

Northeast, Midwest, South, West¹

Percentage Hispanic/Latino in Block

≥ 71%, [20,71)%, < 20%¹

Percentage Owner Occupied in Block

≥ 50%, [10-50)%, < 10%¹

Age Category * Gender²

Male 12-17,¹ Male 18-25,¹ Male 26-34, Male 35-49, Male 50+,¹ Female 12-17,¹ Female 18-25,¹ Female 26-34,¹ Female 35-49,¹ Female 50+¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Age * Race/Hispanic Recode (4 Levels)

Age * Not Hispanic/Latino and White Only,¹ Age * Not Hispanic/Latino and Black/African American Only, Age * Not Hispanic/Latino and Other, Age * Hispanic/Latino

Age Squared * Race/Hispanic Recode (4 Levels)

Age Squared * Not Hispanic/Latino and White Only,¹ Age Squared * Not Hispanic/Latino and Black/African American Only, Age Squared * Not Hispanic/Latino and Other, Age Squared * Hispanic/Latino

Other Family Members in Household

No, Yes¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.7 Income Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Race/Hispanic Recode (2 Levels)

Not Hispanic/Latino and White Only, Hispanic/Latino or Not White Only¹

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only, Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other,¹ Hispanic/Latino

Exhibit D.7 Income Variables (continued)

Education Level (2 Levels)

Not a College Graduate, College Graduate¹

Education Level (3 Levels)

Less Than High School, High School Graduate/Some College, College Graduate¹

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (2 Levels)

Full/Part Time, Unemployed/Other Employment¹

Employment Status (3 Levels)

Full Time, Part Time, Unemployed/Other Employment¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (3 Levels)

Married, Was Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Population Density (2 Levels)

Segment Not in CBSA or Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Census Region

Northeast, Midwest, South, West¹

Percentage Hispanic/Latino in Block (2 Levels)

≥ 20%, < 20%¹

Percentage Hispanic/Latino in Block (3 Levels)

≥ 71%, [20,71)%, < 20%¹

Percentage Non-Hispanic/Latino Black/African American in Block (2 Levels)

≥ 10%, < 10%¹

Percentage Non-Hispanic/Latino Black/African American in Block (3 Levels)

≥ 50%, [10,50)%, < 10%¹

Percentage Owner Occupied in Block (2 Levels)

≥ 10%, < 10%¹

Percentage Owner Occupied in Block (3 Levels)

≥ 50%, [10-50)%, < 10%¹

Gender * Race/Hispanic Recode (2 Levels)

Male Not Hispanic/Latino and White Only, Male Hispanic/Latino or Not White Only,¹ Female Not Hispanic/Latino and White Only,¹ Female Hispanic/Latino or Not White Only¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only, Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other,¹ Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Age * Race/Hispanic Recode (2 Levels)

Age * Not Hispanic/Latino and White Only,¹ Age * Hispanic/Latino or Not White Only

Age * Race/Hispanic Recode (4 Levels)

Age * Not Hispanic/Latino and White Only, Age * Not Hispanic/Latino and Black/African American Only, Age * Not Hispanic/Latino and Other,¹ Age * Hispanic/Latino

Age Squared * Race/Hispanic Recode (2 Levels)

Age Squared * Not Hispanic/Latino and White Only,¹ Age Squared * Hispanic/Latino or Not White Only

Exhibit D.7 Income Variables (continued)

Age Squared * Race/Hispanic Recode (4 Levels)

Age Squared * Not Hispanic/Latino and White Only, Age Squared * Not Hispanic/Latino and Black/African American Only, Age Squared * Not Hispanic/Latino and Other,¹ Age Squared * Hispanic/Latino

Age Cubed * Race/Hispanic Recode (2 Levels)

Age Cubed * Not Hispanic/Latino and White Only,¹ Age Cubed * Hispanic/Latino or Not White Only

Age Cubed * Race/Hispanic Recode (4 Levels)

Age Cubed * Not Hispanic/Latino and White Only, Age Cubed * Not Hispanic/Latino and Black/African American Only, Age Cubed * Not Hispanic/Latino and Other,¹ Age Cubed * Hispanic/Latino

Intermediate Family Received Food Stamps

Yes, No¹

Family Received Food Stamps

No, Yes¹

Intermediate Family Received Public Assistance

Yes, No¹

Family Received Public Assistance

No, Yes¹

Intermediate Family Received Social Security or Railroad Payments

Yes, No¹

Family Received Social Security or Railroad Payments

No, Yes¹

Intermediate Family Received Supplemental Security Income

Yes, No¹

Family Received Supplemental Security Income

No, Yes¹

Intermediate Family Received Welfare/Job Placement/Childcare Services

Yes, No¹

Family Received Welfare/Job Placement/Childcare Services

No, Yes¹

Total Family Income > or < \$20,000

< \$20,000, ≥ \$20,000¹

Income State Rank (2 Levels)

Low/Middle Income State Rank (low/middle tertile), High Income State Rank (high tertile)¹

Income State Rank (3 Levels)

Low State Income Rank (low tertile), Middle State Income Rank (middle tertile), High State Income Rank (high tertile)¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.8 Health Insurance Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender

Male, Female¹

Race/Hispanic Recode (4 Levels)

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Education Level (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (4 Levels)

Full Time, Part Time, Unemployed, Other Employment¹

Exhibit D.8 Health Insurance Variables (continued)

Marital Status (2 Levels)

Married, Never Been Married¹

Marital Status (4 Levels)

Married, Widowed, Divorced/Separated, Never Been Married¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Percentage Owner Occupied in Block

≥ 50%, [10-50)%, < 10%¹

Gender * Race/Hispanic Recode (4 Levels)

Male Not Hispanic/Latino and White Only,¹ Male Not Hispanic/Latino and Black/African American Only, Male Not Hispanic/Latino and Other, Male Hispanic/Latino, Female Not Hispanic/Latino and White Only,¹ Female Not Hispanic/Latino and Black/African American Only,¹ Female Not Hispanic/Latino and Other,¹ Female Hispanic/Latino¹

Age * Race/Hispanic Recode (4 Levels)

Age * Not Hispanic/Latino and White Only,¹ Age * Not Hispanic/Latino and Black/African American Only, Age * Not Hispanic/Latino and Other, Age * Hispanic/Latino

Age Squared * Race/Hispanic Recode (4 Levels)

Age Squared * Not Hispanic/Latino and White Only,¹ Age Squared * Not Hispanic/Latino and Black/African American Only, Age Squared * Not Hispanic/Latino and Other, Age Squared * Hispanic/Latino

Other Family Members in Household

No, Yes¹

Family Participation in Government Assistance Programs

No,¹ Yes

Lifetime Military Service

No,¹ Yes

Family Received Social Security or Railroad Payments

Yes, No¹

Household Size (2 Levels)

At Least Four People in Household, Fewer Than Four People in Household¹

Intermediate Medicare Coverage

Yes, No¹

Intermediate CHAMPUS Coverage

Yes, No¹

Intermediate Medicaid/CHIP Coverage

Yes, No¹

Family Income Recode

Income Less Than \$20,000, Income \$20,000-\$49,999, Income \$50,000-\$74,999, Income \$75,000 or More¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Exhibit D.9 Pair Relationship Variables

Interview Mode (2 Levels)

Web, Field¹

Quarter (4 Levels)

Quarter 1, Quarter 2, Quarter 3, Quarter 4¹

Gender - Non-Pair

Male,¹ Female

Gender (Younger)

Female, Male¹

Gender (Older)

Female, Male¹

Exhibit D.9 Pair Relationship Variables (continued)

Age Category - Non-Pair

12-14, 15-17, 18-20, 21-25, 26-34, 35-49, 50+, Unknown¹

Age Category (Older)

12-14, 15-17, 18-20, 21-25, 26-34, 35-49, 50+, Unknown¹

Race/Hispanic Recode (4 Levels) - Non-Pair

Not Hispanic/Latino and White Only,¹ Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and Other, Hispanic/Latino

Race/Hispanic Recode (Older) (4 Levels)

Hispanic/Latino, Not Hispanic/Latino and Other, Not Hispanic/Latino and Black/African American Only, Not Hispanic/Latino and White Only¹

Education Level (4 Levels) - Non-Pair

Less Than High School, High School Graduate, Some College, College Graduate¹

Education Level (Younger) (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Education Level (Older) (4 Levels)

Less Than High School, High School Graduate, Some College, College Graduate¹

Employment Status (4 Levels) - Non-Pair

Full Time, Part Time, Unemployed, Other Employment¹

Employment Status (Younger)

Full Time, Part Time, Unemployed, Other Employment¹

Employment Status (Older)

Full Time, Part Time, Unemployed, Other Employment¹

Marital Status (2 Levels) - Non-Pair

Married, Widowed or Divorced, Never Been Married¹

Marital Status (Older) (2 Levels)

Married, Widowed or Divorced, Never Been Married¹

Marital Status (Younger) (4 Levels)

Married, Widowed, Divorced, Never Been Married¹

Marital Status (Older) (4 Levels)

Married, Widowed, Divorced, Never Been Married¹

Population Density (3 Levels)

Segment Not in CBSA, Segment in CBSA with Fewer Than 1 Million People, Segment in CBSA with 1 Million or More People¹

Census Region

Northeast, Midwest, South, West¹

Percentage Black/African American in Block

≥ 40%, [10,40)%, < 10%¹

Percentage Hispanic/Latino in Block

≥ 71%, [20,71)%, < 20%¹

Percentage Owner Occupied in Block

≥ 50%, [10,50)%, < 10%¹

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area.

¹ This is the reference level for this variable, against which effects of other factor levels are measured.

Appendix E: Hot-Deck/Collocated Stochastic Imputation Procedure Summaries

Tables

E.1	Likeness Constraints for Marital Status	E-3
E.2	Constraints and Portion of the Predictive Mean Vector for Marital Status	E-3
E.3	Logical Constraints for Race	E-3
E.4	Likeness Constraints for Race	E-3
E.5	Constraints and Portion of the Predictive Mean Vector for Race	E-4
E.6	Likeness Constraints for Hispanic/Latino Origin	E-6
E.7	Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Origin	E-6
E.8	Likeness Constraints for Hispanic/Latino Group	E-6
E.9	Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Group	E-6
E.10	Likeness Constraints for Education Level	E-7
E.11	Constraints and Portion of the Predictive Mean Vector for Education Level	E-7
E.12	Logical Constraints for Employment Status	E-7
E.13	Likeness Constraints for Employment Status	E-7
E.14	Constraints and Portion of the Predictive Mean Vector for Employment Status	E-8
E.15	Likeness Constraints for Indicator of Birth in the United States	E-8
E.16	Constraints and Portion of the Predictive Mean Vector for Indicator of Birth in the United States	E-8
E.17	Logical Constraints for Age of Entry in the United States	E-8
E.18	Likeness Constraints for Age of Entry in the United States	E-9
E.19	Constraints and Portion of the Predictive Mean Vector for Age of Entry in the United States	E-9
E.20	Elements of Full Predictive Mean Vector for Lifetime Substance Use	E-10
E.21	Likeness Constraints for Lifetime Substance Use	E-10
E.22	Constraints and Portion of the Predictive Mean Vector for Lifetime Substance Use	E-11
E.23	Likeness Constraints for Cigarette Ever Daily Used	E-11
E.24	Constraints and Portion of the Predictive Mean Vector for Cigarette Ever Daily Used	E-11
E.25	Logical Constraints for Cigarette Recency and Frequency	E-12
E.26	Likeness Constraints for Cigarette Recency and Frequency	E-12
E.27	Constraints and Portion of the Predictive Mean Vector for Cigarette Recency and Frequency	E-13
E.28	Logical Constraints for Cigar and Smokeless Tobacco Recency and Frequency	E-14
E.29	Likeness Constraints for Cigar and Smokeless Tobacco Recency and Frequency	E-14
E.30	Constraints and Portion of the Predictive Mean Vector for Cigar Recency and Frequency	E-15
E.31	Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency	E-16

E.32	Likeness Constraints for Pipe Recency and Frequency.....	E-17
E.33	Constraints and Portion of the Predictive Mean Vector for Pipe Recency and Frequency	E-17
E.34	Logical Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency	E-17
E.35	Likeness Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency	E-19
E.36	Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency	E-20
E.37	Constraints and Portion of the Predictive Mean Vector for 30-Day Binge Alcohol Frequency	E-20
E.38	Constraints and Portion of the Predictive Mean Vector for Marijuana Recency and Frequency	E-21
E.39	Logical Constraints for Ways of Marijuana Use	E-22
E.40	Likeness Constraints for Ways of Marijuana Use.....	E-22
E.41	Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Year	E-23
E.42	Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Month.....	E-26
E.43	Constraints and Portion of the Predictive Mean Vector for Heroin Recency and Frequency	E-29
E.44	Logical Constraints for Cocaine Recency and Frequency.....	E-30
E.45	Likeness Constraints for Cocaine Recency and Frequency	E-32
E.46	Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency	E-33
E.47	Logical Constraints for Age at First Use, Univariate Assignment	E-37
E.48	Likeness Constraints for Age at First Use, Univariate Assignment.....	E-37
E.49	Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Use	E-38
E.50	Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Daily Use	E-38
E.51	Constraints and Portion of the Predictive Mean Vector for Cigar Age at First Use.....	E-38
E.52	Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Age at First Use	E-39
E.53	Constraints and Portion of the Predictive Mean Vector for Alcohol Age at First Use....	E-39
E.54	Constraints and Portion of the Predictive Mean Vector for Marijuana Age at First Use	E-39
E.55	Constraints and Portion of the Predictive Mean Vector for Heroin Age at First Use.....	E-40
E.56	Logical Constraints for Age at First Use, Multivariate Assignment.....	E-40
E.57	Likeness Constraints for Age at First Use, Multivariate Assignment.....	E-41
E.58	Constraints and Portion of the Predictive Mean Vector for Cocaine Age at First Use	E-41

E.59	Imputation Details for Inhalants Lifetime Use	E-41
E.60	Imputation Details for Methamphetamine Lifetime Use	E-42
E.61	Imputation Details for Hallucinogens Lifetime Use	E-42
E.62	Elements of Full Predictive Mean Vector for Hallucinogen Child Lifetime Drug Use.....	E-42
E.63	Logical Constraints for Hallucinogen Child Lifetime Drug Use.....	E-43
E.64	Likeness Constraints for Hallucinogen Child Lifetime Drug Use	E-43
E.65	Constraints and Portion of the Predictive Mean Vector for Hallucinogen Child Lifetime Drug Use.....	E-43
E.66	Imputation Details for Lifetime Vaping of Nicotine or Tobacco	E-43
E.67	Imputation Details for Lifetime Flavoring Vaping	E-43
E.68	Imputation Details for Illegally Made Fentanyl Lifetime Use.....	E-44
E.69	Imputation Details for Kratom Lifetime Use	E-44
E.70	Imputation Details for Synthetic Marijuana Lifetime Use	E-44
E.71	Imputation Details for Synthetic Stimulants Lifetime Use	E-44
E.72	Imputation Details for CBD or Hemp Lifetime Use	E-44
E.73	Imputation Details for Sedatives Recency of Any Use	E-45
E.74	Imputation Details for Stimulants Recency of Any Use	E-45
E.75	Imputation Details for Tranquilizers Recency of Any Use.....	E-46
E.76	Imputation Details for Pain Relievers Recency of Any Use	E-46
E.77	Imputation Details for OxyContin Recency of Any Use	E-46
E.78	Imputation Details for Prescription Fentanyl Recency of Any Use	E-47
E.79	Imputation Details for Benzodiazepines Recency of Any Use	E-47
E.80	Imputation Details for Inhalants Recency of Use	E-47
E.81	Imputation Details for Methamphetamine Recency of Use.....	E-48
E.82	Imputation Details for Hallucinogens Recency of Use.....	E-48
E.83	Imputation Details for LSD Recency of Use.....	E-49
E.84	Imputation Details for PCP Recency of Use	E-50
E.85	Imputation Details for Ecstasy Recency of Use.....	E-51
E.86	Imputation Details for Ketamine Recency of Use	E-52
E.87	Imputation Details for AMT/DMT/Foxy Recency of Use	E-53
E.88	Imputation Details for <i>Salvia divinorum</i> Recency of Use	E-54
E.89	Imputation Details for Sedatives Recency of Misuse	E-55
E.90	Imputation Details for Stimulants Recency of Misuse	E-56
E.91	Imputation Details for Tranquilizers Recency of Misuse.....	E-57
E.92	Imputation Details for Pain Relievers Recency of Misuse	E-58
E.93	Imputation Details for OxyContin Past Year Recency of Misuse.....	E-58
E.94	Imputation Details for Prescription Fentanyl Past Year Recency of Misuse	E-59

E.95	<u>Imputation Details for Benzodiazepines Past Year Recency of Misuse.....</u>	<u>E-59</u>
E.96	<u>Imputation Details for Nicotine or Tobacco Vaping Recency of Use.....</u>	<u>E-60</u>
E.97	<u>Imputation Details for Flavoring Vaping Recency of Use.....</u>	<u>E-60</u>
E.98	<u>Imputation Details for CBD or Hemp Recency of Use.....</u>	<u>E-61</u>
E.99	<u>Imputation Details for Illegally Made Fentanyl Recency of Use</u>	<u>E-61</u>
E.100	<u>Imputation Details for Kratom Recency of Use</u>	<u>E-62</u>
E.101	<u>Imputation Details for Synthetic Marijuana Recency of Use.....</u>	<u>E-62</u>
E.102	<u>Imputation Details for Synthetic Stimulants Recency of Use</u>	<u>E-63</u>
E.103	<u>Imputation Details for GHB Recency of Use</u>	<u>E-63</u>
E.104	<u>Imputation Details for Nonprescription Cough or Cold Medicine Recency of Use.....</u>	<u>E-64</u>
E.105	<u>Logical Constraints for Inhalants and Methamphetamine 12-Month Frequency.....</u>	<u>E-64</u>
E.106	<u>Likeness Constraints for Inhalants and Methamphetamine 12-Month Frequency</u>	<u>E-64</u>
E.107	<u>Constraints and Portion of the Predictive Mean Vector for Inhalants 12-Month Frequency.....</u>	<u>E-65</u>
E.108	<u>Constraints and Portion of the Predictive Mean Vector for Methamphetamine 12-Month Frequency.....</u>	<u>E-65</u>
E.109	<u>Logical Constraints for Hallucinogens 12-Month Frequency</u>	<u>E-66</u>
E.110	<u>Likeness Constraints for Hallucinogens 12-Month Frequency</u>	<u>E-66</u>
E.111	<u>Constraints and Portion of the Predictive Mean Vector for Hallucinogens 12-Month Frequency.....</u>	<u>E-67</u>
E.112	<u>Logical Constraints for Nicotine Vaping, Inhalants, Methamphetamine, Hallucinogens, Sedatives, Stimulants, Tranquilizers, and Pain Relievers 30-Day Frequency</u>	<u>E-68</u>
E.113	<u>Likeness Constraints for Inhalants, Methamphetamine, Hallucinogens, Sedatives, Stimulants, Tranquilizers, and Pain Relievers 30-Day Frequency</u>	<u>E-68</u>
E.114	<u>Constraints and Portion of the Predictive Mean Vector for Nicotine Vaping 30-Day Frequency</u>	<u>E-69</u>
E.115	<u>Constraints and Portion of the Predictive Mean Vector for Inhalants 30-Day Frequency</u>	<u>E-69</u>
E.116	<u>Constraints and Portion of the Predictive Mean Vector for Methamphetamine 30-Day Frequency</u>	<u>E-70</u>
E.117	<u>Constraints and Portion of the Predictive Mean Vector for Hallucinogens 30-Day Frequency</u>	<u>E-70</u>
E.118	<u>Constraints and Portion of the Predictive Mean Vector for Sedatives 30-Day Frequency</u>	<u>E-71</u>
E.119	<u>Constraints and Portion of the Predictive Mean Vector for Stimulants 30-Day Frequency</u>	<u>E-71</u>
E.120	<u>Constraints and Portion of the Predictive Mean Vector for Tranquilizers 30-Day Frequency</u>	<u>E-72</u>

E.121	<u>Constraints and Portion of the Predictive Mean Vector for Pain Relievers 30-Day Frequency</u>	<u>E-72</u>
E.122	<u>Imputation Details for Past Year Initiation of Misuse for Sedatives.....</u>	<u>E-73</u>
E.123	<u>Imputation Details for Past Year Initiation of Misuse for Stimulants.....</u>	<u>E-73</u>
E.124	<u>Imputation Details for Past Year Initiation of Misuse for Tranquilizers</u>	<u>E-73</u>
E.125	<u>Imputation Details for Past Year Initiation of Misuse for Pain Relievers.....</u>	<u>E-74</u>
E.126	<u>Logical Constraints for modPMN Age at First Use, Univariate.....</u>	<u>E-74</u>
E.127	<u>Likeness Constraints for modPMN Age at First Use, Univariate.....</u>	<u>E-75</u>
E.128	<u>Constraints and Portion of the Predictive Mean Vector for Nicotine Vaping Age at First Use</u>	<u>E-76</u>
E.129	<u>Constraints and Portion of the Predictive Mean Vector for Inhalants Age at First Use</u>	<u>E-76</u>
E.130	<u>Constraints and Portion of the Predictive Mean Vector for Methamphetamine Age at First Use</u>	<u>77</u>
E.131	<u>Constraints and Portion of the Predictive Mean Vector for Hallucinogens Age at First Use</u>	<u>E-77</u>
E.132	<u>Constraints and Portion of the Predictive Mean Vector for LSD Age at First Use</u>	<u>E-78</u>
E.133	<u>Constraints and Portion of the Predictive Mean Vector for PCP Age at First Use</u>	<u>E-79</u>
E.134	<u>Constraints and Portion of the Predictive Mean Vector for Ecstasy Age at First Use</u>	<u>E-80</u>
E.135	<u>Imputation Details for Age of First Misuse for Sedatives.....</u>	<u>E-81</u>
E.136	<u>Imputation Details for Age of First Misuse for Stimulants</u>	<u>E-81</u>
E.137	<u>Imputation Details for Age of First Misuse for Tranquilizers</u>	<u>E-81</u>
E.138	<u>Imputation Details for Age of First Misuse for Pain Relievers</u>	<u>E-82</u>
E.139	<u>Imputation Details for Past Year Alcohol Use Disorder: DSM-5.....</u>	<u>E-82</u>
E.140	<u>Imputation Details for Past Year Alcohol Use Disorder Severity: DSM-5</u>	<u>E-82</u>
E.141	<u>Imputation Details for Past Year Marijuana Use Disorder: DSM-5.....</u>	<u>E-83</u>
E.142	<u>Imputation Details for Past Year Marijuana Use Disorder Severity: DSM-5</u>	<u>E-83</u>
E.143	<u>Imputation Details for Past Year Cocaine Use Disorder: DSM-5.....</u>	<u>E-83</u>
E.144	<u>Imputation Details for Past Year Cocaine Use Disorder Severity: DSM-5</u>	<u>E-84</u>
E.145	<u>Imputation Details for Past Year Methamphetamine Use Disorder: DSM-5</u>	<u>E-84</u>
E.146	<u>Imputation Details for Past Year Methamphetamine Use Disorder Severity: DSM-5</u>	<u>E-85</u>
E.147	<u>Imputation Details for Past Year Stimulant Use Disorder: DSM-5</u>	<u>E-85</u>
E.148	<u>Imputation Details for Past Year Stimulant Use Disorder Severity: DSM-5.....</u>	<u>E-86</u>
E.149	<u>Imputation Details for Past Year Heroin Use Disorder: DSM-5</u>	<u>E-86</u>
E.150	<u>Imputation Details for Past Year Heroin Use Disorder Severity: DSM-5</u>	<u>E-87</u>
E.151	<u>Imputation Details for Past Year Pain Reliever Use Disorder: DSM-5</u>	<u>E-87</u>
E.152	<u>Imputation Details for Past Year Pain Reliever Use Disorder Severity: DSM-5.....</u>	<u>E-88</u>

E.153	Imputation Details for Past Year Hallucinogen Use Disorder: DSM-5	E-88
E.154	Imputation Details for Past Year Hallucinogen Use Disorder Severity: DSM-5	E-89
E.155	Imputation Details for Past Year Inhalant Use Disorder: DSM-5	E-89
E.156	Imputation Details for Past Year Inhalant Use Disorder Severity: DSM-5	E-90
E.157	Imputation Details for Past Year Tranquilizer Use Disorder: DSM-5	E-90
E.158	Imputation Details for Past Year Tranquilizer Use Disorder Severity: DSM-5	E-91
E.159	Imputation Details for Past Year Sedative Use Disorder: DSM-5	E-92
E.160	Imputation Details for Past Year Sedative Use Disorder Severity: DSM-5	E-93
E.161	Imputation Details for Inpatient Substance Use Treatment in a Hospital.....	E-93
E.162	Imputation Details for Inpatient Substance Use Treatment in a Residential Drug or Alcohol Rehab or Treatment Center	E-94
E.163	Imputation Details for Inpatient Substance Use Treatment in a Residential Mental Health Treatment Center	E-94
E.164	Imputation Details for Substance Use Treatment in Some Other Place as an Inpatient.....	E-94
E.165	Imputation Details for Outpatient Substance Use Treatment at a Drug or Alcohol Rehab Treatment Center.....	E-94
E.166	Imputation Details for Outpatient Substance Use Treatment at a Mental Health Treatment Center	E-95
E.167	Imputation Details for Outpatient Substance Use Treatment at an Office of a Therapist, Psychologist, Psychiatrist, or Substance Use Treatment Professional	E-95
E.168	Imputation Details for Outpatient Substance Use Treatment at a General Medical Clinic or Doctor’s Office.....	E-95
E.169	Imputation Details for Outpatient Substance Use Treatment at a Hospital	E-95
E.170	Imputation Details for Outpatient Substance Use Treatment at a School Health or Counseling Center	E-96
E.171	Imputation Details for Substance Use Treatment at Some Other Place as an Outpatient	E-96
E.172	Imputation Details for Substance Use Treatment as Part of a Support Group.....	E-96
E.173	Imputation Details for Substance Use Treatment in a Hospital Emergency Department	E-96
E.174	Imputation Details for Medication-Assisted Treatment for Alcohol Use.....	E-96
E.175	Imputation Details for Medication-Assisted Treatment for Opioid Use.....	E-97
E.176	Imputation Details for Use of Telehealth Substance Use Treatment	E-97
E.177	Imputation Details for Substance Use Treatment in Prison, Jail, or Juvenile Detention Center	E-97
E.178	Imputation Details for Use of Detoxification or Withdrawal Support Services	E-97
E.179	Imputation Details for Substance Use Treatment from a Peer Support Specialist or Recovery Coach	E-97
E.180	Likeness Constraints for Kessler-6 (K6) Psychological Distress Scale	E-98

E.181	<u>Constraints and Portion of the Predictive Mean Vector for Past Month Kessler-6 (K6) Psychological Distress Scale</u>	<u>E-99</u>
E.182	<u>Constraints and Portion of the Predictive Mean Vector for Worst Month in Past Year Kessler-6 (K6) Psychological Distress Scale</u>	<u>E-100</u>
E.183	<u>Imputation Details for Kessler-6 (K6) Psychological Distress Scale, Month in Past Year Worse than Past 30 Days</u>	<u>E-101</u>
E.184	<u>Logical Constraints for WHODAS.....</u>	<u>E-101</u>
E.185	<u>Likeness Constraints for WHODAS</u>	<u>E-102</u>
E.186	<u>Constraints and Portion of the Predictive Mean Vector for WHODAS</u>	<u>E-103</u>
E.187	<u>Imputation Details for Lifetime MDE.....</u>	<u>E-104</u>
E.188	<u>Imputation Details for Past Year MDE.....</u>	<u>E-104</u>
E.189	<u>Imputation Details for Past Year MDE with Impairment.....</u>	<u>E-104</u>
E.190	<u>Imputation Details for Suicidal Thoughts</u>	<u>E-104</u>
E.191	<u>Imputation Details for Suicidal Plans</u>	<u>E-105</u>
E.192	<u>Imputation Details for Suicidal Attempts.....</u>	<u>E-105</u>
E.193	<u>Imputation Details for Suicidal Thoughts due to Coronavirus Disease 2019 (COVID-19)</u>	<u>E-105</u>
E.194	<u>Imputation Details for Suicidal Plans due to Coronavirus Disease 2019 (COVID-19)</u>	<u>E-105</u>
E.195	<u>Imputation Details for Suicidal Attempts due to Coronavirus Disease 2019 (COVID-19)</u>	<u>E-106</u>
E.196	<u>Imputation Details for Receive Medical Attention for Suicidal Attempt</u>	<u>E-106</u>
E.197	<u>Imputation Details for Overnight or Longer Hospital Stay for Suicidal Attempt</u>	<u>E-106</u>
E.198	<u>Imputation Details for Inpatient Mental Health Treatment in a Hospital.....</u>	<u>E-107</u>
E.199	<u>Imputation Details for Inpatient Mental Health Treatment in a Residential Mental Health Treatment Center</u>	<u>E-107</u>
E.200	<u>Imputation Details for Inpatient Mental Health Treatment in a Residential Drug or Alcohol Rehab or Treatment Center</u>	<u>E-107</u>
E.201	<u>Imputation Details for Mental Health Treatment in Some Other Place as an Inpatient.....</u>	<u>E-107</u>
E.202	<u>Imputation Details for Outpatient Mental Health Treatment at a Mental Health Treatment Center.....</u>	<u>E-108</u>
E.203	<u>Imputation Details for Outpatient Mental Health Treatment at a Drug or Alcohol Rehab Treatment Center.....</u>	<u>E-108</u>
E.204	<u>Imputation Details for Outpatient Mental Health Treatment at an Office of a Therapist, Psychologist, Psychiatrist, or Mental Health Treatment Professional.....</u>	<u>E-108</u>
E.205	<u>Imputation Details for Outpatient Mental Health Treatment at a General Medical Clinic or Doctor’s Office.....</u>	<u>E-108</u>
E.206	<u>Imputation Details for Outpatient Mental Health Treatment at a Hospital</u>	<u>E-109</u>

E.207	<u>Imputation Details for Outpatient Mental Health Treatment at a School Health or Counseling Center</u>	<u>E-109</u>
E.208	<u>Imputation Details for Mental Health Treatment at Some Other Place as an Outpatient</u>	<u>E-109</u>
E.209	<u>Imputation Details for Mental Health Treatment as Part of a Support Group</u>	<u>E-109</u>
E.210	<u>Imputation Details for Mental Health Treatment in a Hospital Emergency Department</u>	<u>E-109</u>
E.211	<u>Imputation Details for Use of Prescription Medication for Mental Health Treatment</u>	<u>E-110</u>
E.212	<u>Imputation Details for Use of Telehealth Mental Health Treatment</u>	<u>E-110</u>
E.213	<u>Imputation Details for Mental Health Treatment in Prison, Jail, or Juvenile Detention Center</u>	<u>E-110</u>
E.214	<u>Imputation Details for Mental Health Treatment from a Peer Support Specialist or Recovery Coach</u>	<u>E-110</u>
E.215	<u>Logical Constraints for Number of Rostered People</u>	<u>E-111</u>
E.216	<u>Likeness Constraints for Number of Rostered People</u>	<u>E-111</u>
E.217	<u>Constraints and Portion of the Predictive Mean Vector for Number of Rostered People</u>	<u>E-111</u>
E.218	<u>Logical Constraints for Number of Children Younger than 18</u>	<u>E-111</u>
E.219	<u>Likeness Constraints for Number of Children Younger than 18</u>	<u>E-111</u>
E.220	<u>Constraints and Portion of the Predictive Mean Vector for Number of Children Younger than 18</u>	<u>E-111</u>
E.221	<u>Logical Constraints for Number of People Aged 65 or Older</u>	<u>E-111</u>
E.222	<u>Likeness Constraints for Number of People Aged 65 or Older</u>	<u>E-112</u>
E.223	<u>Constraints and Portion of the Predictive Mean Vector for Number of People Aged 65 or Older</u>	<u>E-112</u>
E.224	<u>Likeness Constraints for Indicator of Whether the Respondent Has Family Members in Household</u>	<u>E-112</u>
E.225	<u>Constraints and Portion of the Predictive Mean Vector for Indicator of Whether the Respondent Has Family Members in Household</u>	<u>E-112</u>
E.226	<u>Logical Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)</u>	<u>E-112</u>
E.227	<u>Likeness Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)</u>	<u>E-113</u>
E.228	<u>Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (Including Foster Relationships)</u>	<u>E-113</u>
E.229	<u>Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)</u>	<u>E-113</u>
E.230	<u>Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)</u>	<u>E-113</u>

E.231	<u>Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)</u>	<u>E-114</u>
E.232	<u>Logical Constraints for Source of Income.....</u>	<u>E-114</u>
E.233	<u>Likeness Constraints for Source of Income</u>	<u>E-115</u>
E.234	<u>Constraints and Portion of the Predictive Mean Vector for Source of Income</u>	<u>E-116</u>
E.235	<u>Logical Constraints for Finer Categories of Income</u>	<u>E-117</u>
E.236	<u>Likeness Constraints for Finer Categories of Income.....</u>	<u>E-117</u>
E.237	<u>Constraints and Portion of the Predictive Mean Vector for Finer Categories of Income</u>	<u>E-117</u>
E.238	<u>Likeness Constraints for Health Insurance (MPMN).....</u>	<u>E-118</u>
E.239	<u>Constraints and Portion of the Predictive Mean Vector for Health Insurance (MPMN).....</u>	<u>E-119</u>
E.240	<u>Constraints for Health Insurance (UPMN), Any Other Health Insurance</u>	<u>E-121</u>
E.241	<u>Constraints and Portion of the Predictive Mean Vector for Health Insurance (UPMN), Any Other Health Insurance</u>	<u>E-121</u>
E.242	<u>Variables Used in Constraints for Stage One, Pair Relationship</u>	<u>E-122</u>
E.243	<u>Logical Constraints for Stage One, Pair Relationship</u>	<u>E-123</u>
E.244	<u>Likeness Constraints for Stage One, Pair Relationship</u>	<u>E-124</u>
E.245	<u>Constraints and Portion of the Predictive Mean Vector for Stage One, Pair Relationship.....</u>	<u>E-125</u>
E.246	<u>Likeness Constraints for Parent-Child, Child Focus Multiplicity Counts.....</u>	<u>E-126</u>
E.247	<u>Constraints and Portion of the Predictive Mean Vector for Parent-Child, Child Focus Multiplicity Counts.....</u>	<u>E-126</u>
E.248	<u>Logical Constraints for Parent-Child, Parent Focus Multiplicity Counts.....</u>	<u>E-126</u>
E.249	<u>Likeness Constraints for Parent-Child, Parent Focus Multiplicity Counts</u>	<u>E-127</u>
E.250	<u>Constraints and Portion of the Predictive Mean Vector for Parent-Child, Parent Focus Multiplicity Counts.....</u>	<u>E-128</u>
E.251	<u>Logical Constraints for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts</u>	<u>E-129</u>
E.252	<u>Likeness Constraints for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts</u>	<u>E-129</u>
E.253	<u>Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts.....</u>	<u>E-129</u>
E.254	<u>Logical Constraints for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts</u>	<u>E-129</u>
E.255	<u>Likeness Constraints for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts</u>	<u>E-130</u>
E.256	<u>Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts</u>	<u>E-130</u>

E.257	<u>Logical Constraints for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts</u>	<u>E-130</u>
E.258	<u>Likeness Constraints for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts</u>	<u>E-130</u>
E.259	<u>Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts.....</u>	<u>E-131</u>
E.260	<u>Logical Constraints for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts</u>	<u>E-131</u>
E.261	<u>Likeness Constraints for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts</u>	<u>E-131</u>
E.262	<u>Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts</u>	<u>E-131</u>
E.263	<u>Logical Constraints for Parent-Child Household Counts</u>	<u>E-132</u>
E.264	<u>Likeness Constraints for Stage Three, Household Counts</u>	<u>E-133</u>
E.265	<u>Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts.....</u>	<u>E-134</u>
E.266	<u>Logical Constraints for Household Sibling Count (12-14/15-17), Older Focus</u>	<u>E-142</u>
E.267	<u>Likeness Constraints for Household Sibling Count (12-14/15-17), Older Focus</u>	<u>E-142</u>
E.268	<u>Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-14/15-17), Older Focus</u>	<u>E-142</u>
E.269	<u>Logical Constraints for Household Sibling Count (12-17/18-25), Older Focus</u>	<u>E-143</u>
E.270	<u>Likeness Constraints for Household Sibling Count (12-17/18-25), Older Focus</u>	<u>E-143</u>
E.271	<u>Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-17/18-25), Older Focus</u>	<u>E-143</u>
E.272	<u>Variables Used in Constraints for the Household Spouse-Spouse Count.....</u>	<u>E-144</u>
E.273	<u>Logical Constraints for the Household Spouse-Spouse Count.....</u>	<u>E-144</u>
E.274	<u>Likeness Constraints for Household Spouse-Spouse Count</u>	<u>E-144</u>
E.275	<u>Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse Count.....</u>	<u>E-145</u>
E.276	<u>Variables Used in Constraints for Household Spouse-Spouse with Children Count....</u>	<u>E-145</u>
E.277	<u>Logical Constraints for Household Spouse-Spouse with Children Count</u>	<u>E-145</u>
E.278	<u>Likeness Constraints for Household Spouse-Spouse with Children Count</u>	<u>E-146</u>
E.279	<u>Constraints and Portion of the Predictive Mean Vector for the Household Spouse-Spouse with Children Count.....</u>	<u>E-147</u>

E.1 Introduction

For the majority of variables that had missing values imputed in the 2023 National Survey on Drug Use and Health (NSDUH), either the predictive mean neighborhood (PMN) or the modified predictive mean neighborhood (modPMN) imputation methodology was used.¹ The imputation step of PMN and modPMN involves using either a hot-deck step or collocated stochastic imputation (CSI) to replace missing values. For variables imputed using a hot-deck step, this appendix summarizes the predictive mean vectors and the constraints applied on the donor pool. For variables imputed using CSI, this appendix includes information on the edited values, the restrictions placed on the imputed values to maintain consistency with the previously imputed variables, and the predictive mean vectors. It is organized by groups of variables requiring imputation: demographics, substance use, household composition (roster), income, health insurance, and pair.

There are three types of tables associated with each variable or set of variables imputed using the hot-deck step: (1) Logical Constraints, (2) Likeness Constraints, and (3) Constraints and Portion of the Predictive Mean Vector.

For variables that do not apply any logical constraints in the hot-deck step, the “Logical Constraints” table is not applicable. The “Constraints and Portion of the Predictive Mean Vector” table specifies the following for each missingness pattern:

1. the number of item nonrespondents in 2023 exhibiting the pattern (“Total Number of Cases”);
2. the set of logical constraints applied to the potential donors (“Logical Constraints”);
3. the elements of the predictive mean vector (“Predictive Mean Vector”) used to calculate the Mahalanobis distance from recipient to potential donor and to restrict the donor set via the delta constraint; and
4. the set of likeness constraints used in each try and the number of item nonrespondents in 2023 who found donors on each try by different age groups (“Likeness Constraints: Number of Cases”).²

In the tables that follow, the phrase “Donor’s predicted means each must be within x percent of recipient’s predicted means” appears in each of the multivariate imputation tables, and the phrase “Donor’s predicted mean must be within x percent of recipient’s predicted mean” appears in each of the univariate imputation tables. In either case, it represents one of the likeness constraints, known as the “delta constraint,” and also defines the neighborhood. When this constraint is loosened, the neighborhood of potential donors is abandoned, and the candidate with the predicted mean closest to the recipient’s (subject to the constraints that are still on the pool of donors) is chosen as the donor.

¹ See Chapter 3 for details on PMN and modPMN imputation.

² If a cell contains the text “No Cases” within “Likeness Constraints: Number of Cases,” this means that there were no cases requiring imputation in that category.

Although statistical imputation of the substance use or income variables could not have proceeded separately within each state because of insufficient pools of donors, the hot-deck step does incorporate information about the state of residence of each respondent. For the substance use variables, in the hot-deck step, respondents were separated into three state usage-level categories for each substance, depending on the response variable of interest. Respondents from states with high usage of a given substance were placed in one category, respondents from medium-usage states were placed in another category, and respondents from low-usage states were placed in a third category. For the income variables, the states were separated into three income groups, depending upon the proportion of families within those states with incomes greater than or equal to \$20,000. As with the substance use variables, respondents from high-income states (by this measure) were placed in one category, respondents from medium-income states were placed in another category, and respondents from low-income states were placed in a third category. In the tables that follow, this variable is identified as the “state rank” for the substance use and income variables. It was used as a likeness constraint, where the set of eligible donors for each recipient was restricted so that donors and recipients were both from states with the same state rank.

E.2 Demographics

[Tables E.1](#) through [E.11](#) present information on the imputation procedures for the interviewer-administered demographic variables: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level. [Tables E.12](#) through [E.19](#) present information for the self-administered demographic variables: employment status, indicator of birth in the United States, and immigrant age of entry into the United States. In several instances, variable names are used without description for the purposes of brevity (see Sections 4.3.2 and 4.3.3 for details). The segment-level variable, SEGID (segment ID), was used only in the likeness constraints for demographic imputation.

As described in the *2023 National Survey on Drug Use and Health (NSDUH) Methodological Resource Book, Section 2: Sample Design Report* (Center for Behavioral Health Statistics and Quality, 2025a) within each state, state sampling (SS) regions were formed, which were further partitioned into clusters of adjacent blocks called “segments.” The segment ID number was a two-letter state abbreviation followed by a two-digit SS region and a two-digit segment identifier, which uniquely identifies each segment. Although the segment identifier was not used as a covariate because of the large number of levels, it was used as a constraint in the hot-deck step of the PMN procedure for race, Hispanicity, education, and employment status, as noted in Chapters 4 and 5.

E.2.1 Marital Status Variables

Table E.1 Likeness Constraints for Marital Status

Constraint #	Likeness Constraint
LikC1	Donor's age must be within 3 years of recipient's age
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.2 Constraints and Portion of the Predictive Mean Vector for Marital Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	89	None	1. M1	1,2: 23	1,2: 23	1,2: 27
				2. M2	1: 7	1: 3	1: 6
				3. M3			

¹ The predictive mean vector components are defined by the following:

1. M1 = P(Married)
2. M2 = P(Widowed)
3. M3 = P(Divorced or separated)

E.2.2 Race Variables

Table E.3 Logical Constraints for Race

Constraint #	Logical Constraint
LogC1	Donor must be Asian, in part or in full
LogC2	Donor must be more than one race
LogC3	Donor must not be White only
LogC4	Donor must be White only or White and American Indian/Alaska Native only
LogC5	Donor must not be American Indian/Alaska Native, in part or in full
LogC6	Donor must be Native Hawaiian/Other Pacific Islander, in part or in full

Table E.4 Likeness Constraints for Race

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC3	If recipient was Hispanic/Latino nonspecific, then donor must be of Hispanic/Latino origin
LikC4	If recipient selected one or more Hispanic/Latino categories, including Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spaniard, then donor's Hispanic/Latino group value must be equal to one of the Hispanic/Latino groups mentioned by recipient
LikC5	Donor must be Mexican (Hispanic/Latino or non-Hispanic/Latino)
LikC6	Donor must be Cuban (Hispanic/Latino or non-Hispanic/Latino)
LikC7	Donor must be Central or South American (Hispanic/Latino or non-Hispanic/Latino)
LikC8	Donor must be Dominican (Hispanic/Latino or non-Hispanic/Latino)
LikC9	Donor must be Spanish (Hispanic/Latino or non-Hispanic/Latino)
LikC10	Donor must be Puerto Rican (Hispanic/Latino or non-Hispanic/Latino)
LikC11	Donor must be Puerto Rican or Cuban (Hispanic/Latino or non-Hispanic/Latino)

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	976	None	1. R1	1-4: 75	1-4: 21	1-4: 29
				2. R2	2-4: 155	2-4: 80	2-4: 67
				3. R3	3,4: 180	3,4: 139	3,4: 230
				4. R4			
				5. R5			
2	Known to be Asian	57	1	1. R1	1-4: 1	1-4: 0	1-4: 1
				2. R2	2-4: 0	2-4: 2	2-4: 6
				3. R3	3,4: 15	3,4: 12	3,4: 20
				4. R4			
				5. R5			
3	Known to be multiple race, but no other information	4	2	1. R1	1-4: 0	1-4: 0	1-4: 0
				2. R2	2-4: 0	2-4: 0	2-4: 0
				3. R3	3,4: 1	3,4: 1	3,4: 2
				4. R4			
				5. R5			
4	Known not to be White, but no other information	4	3	1. R2/(1-R1)	1-4: 0	No Cases	1-4: 0
				2. R3/(1-R1)	2-4: 1		2-4: 1
				3. R4/(1-R1)			3,4: 2
				4. R5/(1-R1)			
5	Known to be White or both White and American Indian/Alaska Native	0	4	1. R1/(1-R2-R3-R4-R5)	No Cases	No Cases	No Cases
6	Known not to be American Indian/Alaska Native, in part or in full	0	5	1. R1/(1-R3)	No Cases	No Cases	No Cases
				2. R2/(1-R3)			
				3. R4/(1-R3)			
				4. R5/(1-R3)			
7	Known to be non-Hispanic Mexican	1	None	1. R1	1-5: 1	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
8	Known to be non-Hispanic Cuban	0	None	1. R1	No Cases	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
9	Known to be non-Hispanic Central or South American	0	None	1. R1	No Cases	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
10	Known to be non-Hispanic Dominican	1	None	1. R1	1-4,8: 0 2-4,8: 0 3,4,8: 1	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
11	Known to be non-Hispanic Spanish	0	None	1. R1	No Cases	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
12	Known to be non-Hispanic Puerto Rican	0	None	1. R1	No Cases	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
13	Known to be non-Hispanic Cuban and Puerto Rican	0	None	1. R1	No Cases	No Cases	No Cases
				2. R2			
				3. R3			
				4. R4			
				5. R5			
14	Known to be Native Hawaiian/Other Pacific Islander	77	6	1. R1	1-4: 1 2-4: 0 3,4: 26	1-4: 1 2-4: 1 3,4: 17	1-4: 0 2-4: 0 3,4: 31
				2. R2			
				3. R3			
				4. R4			
				5. R5			

¹ The predictive mean vector components are defined by the following:

1. R1 = P(White only)
2. R2 = P(Black only)
3. R3 = P(American Indian/Alaska Native only)
4. R4 = P(Asian only)
5. R5 = P(Native Hawaiian/Other Pacific Islander only)

E.2.3 Hispanic/Latino Origin Variables

Table E.6 Likeness Constraints for Hispanic/Latino Origin

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.7 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Origin

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	226	None	1. H1	1,2: 73 2: 108	1,2: 8 2: 11	1,2: 9 2: 17

¹ The predictive mean vector components are defined by the following:
1. H1 = P(Hispanic/Latino origin)

E.2.4 Hispanic/Latino Group Variables

Table E.8 Likeness Constraints for Hispanic/Latino Group

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient had $8 \leq \text{EDHOGRP} \leq 37$, then donor's values for IRRACWH, IRRACEBK, IRRACENA, IRRACEAS, IRRACENH, and IRRACEPI must indicate a subset of the racial categories mentioned by recipient

Table E.9 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Group

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases ²
1	Completely missing	399	None	1. H1 2. H2 3. H3	1-3: 61 2,3: 284 3: 54

¹ The predictive mean vector components are defined by the following:
1. H1 = P(Mexican)
2. H2 = P(Puerto Rican)
3. H3 = P(Central or South American)

² The hot-deck program for the Hispanic/Latino group is not separated into age groups.

E.2.5 Education Variables

Table E.10 Likeness Constraints for Education Level

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = age of recipient

Table E.11 Constraints and Portion of the Predictive Mean Vector for Education Level

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector, by Age Group ¹		Likeness Constraints: Number of Cases, by Age Group		
				12-17	18+	12-17	18-25	26+
1	Missing	90	None	1. E1	1. E1	1-3: 1	1-3: 2	1,2: 3
				2. E2	2. E2	2,3: 6	2,3: 40	2: 33
				3. E3	3. E3	3: 1	3: 4	
				4. E4				

¹ The predictive mean vector components are defined by the following:

12-17 Age Group

1. E1 = P(Less than elementary school)

2. E2 = P(Elementary school)

3. E3 = P(Middle school)

4. E4 = P(Some high school)

18+ Age Group

1. E1 = P(Less than high school)

2. E2 = P(High school)

3. E3 = P(Some college)

E.2.6 Employment Variables

Table E.12 Logical Constraints for Employment Status

Constraint #	Logical Constraint
LogC1	Donor must be employed

Table E.13 Likeness Constraints for Employment Status

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Donor's age must be within 5 years of recipient's age

Table E.14 Constraints and Portion of the Predictive Mean Vector for Employment Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	2,029	None	1. E1	1-3: 13	1-3: 14	1-3: 16
				2. E2	2,3: 119	2,3: 541	2,3: 1,009
				3. E3	3: 16	3: 49	3: 252
2	Known to be employed; part-time vs. full-time status unknown	99	1	1. E1/ (E1+E2)	1-3: 4 2,3: 21	1-3: 7 2,3: 31	1-3: 4 2,3: 32

¹ The predictive mean vector components are defined by the following:

1. E1 = P(Employed full time)
2. E2 = P(Employed part time)
3. E3 = P(Unemployed)

E.2.7 Immigrant Variables

Table E.15 Likeness Constraints for Indicator of Birth in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.16 Constraints and Portion of the Predictive Mean Vector for Indicator of Birth in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	1,638	None	1. B1	1,2: 75 2: 110	1,2: 78 2: 360	1,2: 216 2: 799

¹ The predictive mean vector components are defined by the following:

1. B1 = P(Born in the United States)

Table E.17 Logical Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LogC1	Donor's age of entry must be less than recipient's current age
LogC2	Difference between recipient's current age and donor's age of entry must be less than or equal to 1 if recipient lived in the United States less than a year, or difference must be greater than 1 if recipient lived in the United States more than a year

Table E.18 Likeness Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.19 Constraints and Portion of the Predictive Mean Vector for Age of Entry in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases ²
1	Completely missing	391	1,2	1. PrAgeEntry	1,2: 96 2: 291 None: 3 None*: 1

*A donor could not be found for this case. Date of entry was randomly assigned within bounds and age of entry was calculated based on that value.

¹ The predictive mean vector components are defined by the following:

1. PrAgeEntry = Predicted age of entry

² The hot-deck program for immigrant age of entry is not separated into age groups.

E.3 Substance Use and Substance Use Treatment Variables

Substance use variables were imputed with two methods, PMN and modPMN (see Chapter 3 for details). Cigarettes, cigars, smokeless tobacco, pipes, alcohol, marijuana (including ways of use), heroin, cocaine, and crack are presented in Section E.3.1, PMN. Inhalants, methamphetamine, hallucinogens, hallucinogen child drugs, nicotine vaping, vaping of flavoring, illegally made fentanyl, kratom, synthetic marijuana, synthetic stimulants, CBD or hemp, sedatives, stimulants, tranquilizers, pain relievers, OxyContin, prescription fentanyl, benzodiazepines, GHB, nonprescription cough or cold medications, and substance use treatment are presented in Section E.3.2, modPMN.

Section E.3.1 presents information on the missingness patterns, constraints, and predictive mean vectors applied during PMN imputation procedures for lifetime substance use, recency and frequency of use, and age at first use. [Tables E.20](#) through [E.58](#) contain information on the imputation procedures for substances that continued to be imputed using PMN, with the exception of hallucinogen child drug variables. Although lifetime hallucinogen child drug variables were imputed using PMN methodology, they are included in Section E.3.2 to allow proximity to the parent measure, hallucinogens lifetime.

Section E.3.2 presents information on the imputation procedures applied under the modPMN method. For continuous variables, information on the missingness patterns, constraints, and predictive mean vectors used for hot-deck imputation is presented. For categorical variables, information on the edited values, restrictions from previously imputed variables, and predictive mean vectors used for CSI is presented. [Tables E.59](#) through [E.179](#) contain information by measure on the imputation procedures for substance use and substance use treatment variables that were imputed using modPMN.

E.3.1 PMN

E.3.1.1 Lifetime Substance Use

There were a large number of missingness patterns for lifetime substance use. The usable case rule (Section 2.2.3) required a certain minimum response pattern, but other than that any combination of the lifetime substance use variables may have been missing for a given respondent record.

The probabilities associated with the eight lifetime substance use variables listed in [Table E.20](#) formed the full predictive mean vector. Only the probabilities associated with the lifetime substance use variables for which values were missing were used in the predictive mean vector for each item nonrespondent. The predicted mean for cigarette lifetime use was never calculated because no respondents were missing the lifetime variable for cigarette use.³

Table E.20 Elements of Full Predictive Mean Vector for Lifetime Substance Use

Lifetime Substance Use	Predicted Mean
Heroin Lifetime	P(Lifetime User)
Crack Lifetime	P(Lifetime User Lifetime User of Cocaine), if cocaine lifetime use is known P(Lifetime User of Cocaine) * P(Lifetime User Lifetime User of Cocaine), if cocaine lifetime use is unknown ¹
Cocaine Lifetime	P(Lifetime User)
Marijuana Lifetime	P(Lifetime User)
Alcohol Lifetime	P(Lifetime User)
Pipes Lifetime	P(Lifetime User)
Cigars Lifetime	P(Lifetime User)
Smokeless Tobacco Lifetime	P(Lifetime User)

¹ The predicted mean for individuals missing crack lifetime where cocaine lifetime use is unknown can also be written as the joint probability of P(Lifetime User of Cocaine and Lifetime User of Crack).

Table E.21 Likeness Constraints for Lifetime Substance Use

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Lifetime use of donor = lifetime use of recipient, for each nonmissing lifetime indicator
LikC4	If recipient was missing the lifetime indicator(s) for any member of a family of substances, ¹ then donor's lifetime indicator(s) agreed with recipient's nonmissing lifetime indicator(s) within that family

¹ The cocaine family includes crack.

³ The lifetime variable for cigarette use was never missing because the usable case rule required a nonmissing response to the cigarette gate question.

Table E.22 Constraints and Portion of the Predictive Mean Vector for Lifetime Substance Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing one or more of the lifetime substance use variables	223	None	1. LF for gate question associated with each missing variable	1-3: 129 1,2,4: 5 1,4: 1	1-3: 24 1,2,4: 2 1,4: 5	1-3: 53 1,2,4: 3 1,4: 1

¹ The predictive mean vector components are defined by the following:

1. LF = P(Lifetime user of x), where x is the substance associated with the missing gate question

Table E.23 Likeness Constraints for Cigarette Ever Daily Used

Constraint #	Likeness Constraint
LikC1	Age of donor = age of recipient
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC4	If the recipient was a past year cigarette user, then donor must also be a past year cigarette user
LikC5	If the recipient was a past 3 years but not past year or lifetime but not past 3 years cigarette user, then donor's cigarette recency must match the recipient's cigarette recency
LikC6	If the recipient was a past 3 years but not past year cigarette user, then donor cannot have used cigarettes in the past year
LikC7	If the recipient was a lifetime but not past 3 years cigarette user, then donor cannot have used cigarettes in the past 3 years
LikC8	Age of donor must be greater than or equal to age of recipient

Table E.24 Constraints and Portion of the Predictive Mean Vector for Cigarette Ever Daily Used

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing CIGDLYMO	18	None	1. CIG	1-5: 0 1,3-5: 1	1-5: 6	1-5: 8 1,3-5: 3

¹ The predictive mean vector components are defined by the following:

1. CIG = P(Daily use at some point in lifetime | respondent was a lifetime user but not a current daily user)

E.3.1.2 Recency and Frequency of Substance Use

Many tables in this section abbreviate certain words. "Recency" is an abbreviation for "recency of use," "frequency" or "Freq." is an abbreviation for "frequency of use," and "30-day binge alcohol frequency" or "ALCBNG30D" is an abbreviation for the "number of days in the past 30 days when the respondent consumed five or more alcoholic drinks for males and four or more alcoholic drinks for females."

An empty cell within "Missingness Pattern" indicates that all information was available. An entry of "Missing" in a cell indicates that all information was missing. Other entries for missingness patterns provide the available information, indicating that this information was partially missing. For example, an entry of "Past year" indicates that the individual used the substance of interest in the past year, but the specific recency of past month versus past year not past month must be imputed. However, if the entry is shown in parentheses, then the specific recency is known and no imputation is required. For example, an entry of "(Past year)" indicates that the individual used the substance of interest in the past year and the specific recency of past month versus past year not past month is already known and does not require imputation.

Table E.25 Logical Constraints for Cigarette Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used cigarettes within the past 3 years (a cigarette recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month cigarette user (cigarette recency $\neq$ 1)
LogC3	Donor cannot be a past year cigarette user (cigarette recency $\neq$ 1 or 2)
LogC4	Donor must be a past year cigarette user (cigarette recency = 1 or 2)
LogC5	Donor must be a past month cigarette user (cigarette recency = 1)
LogC6	If recipient was never a daily user of cigarettes (CIGDLYMO = 2), then donor's 30-day cigarette frequency cannot equal 30
LogC7	If recipient's age at first use equals recipient's current age, then donor's 30-day cigarette frequency (1) cannot be greater than the number of days between recipient's interview date and recipient's date of first cigarette use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and recipient's birthday (inclusive)
LogC8	Donor must be a past year but not past month cigarette user or a past 3 years but not past year cigarette user (cigarette recency = 2 or 3)

Table E.26 Likeness Constraints for Cigarette Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.27 Constraints and Portion of the Predictive Mean Vector for Cigarette Recency and Frequency

#	Missingness Pattern ¹	Total Number of Cases	Logical Constraints	Predictive Mean Vector ²	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	6	4,6	1. $R1/(R1+R2)$ 2. $(R1*D)/(R1+R2)$ 3. $R1*(1-D)*PM/(R1+R2)$	No Cases	1,2: 1 1: 4	1,2: 0 1: 1
2	Recency and 30-day frequency missing	6	6	1. R1 2. R2 3. R3 4. $R1*D$ 5. $R1*(1-D)*PM$	No Cases	1,2: 0 1: 2	1,2: 0 1: 4
3	Past month, 30-day frequency missing	9	5-7	1. D 2. PM	1,2: 1 1: 4	1,2: 1	1,2: 3
4	Not past year	98	3,6	1. $R3/(R3+R4)$	1,2: 33 1: 4	1,2: 53	1,2: 8
5	Not past month	44	2,6	1. $R2/(R2+R3+R4)$ 2. $R3/(R2+R3+R4)$	1,2: 10 1: 7	1,2: 20	1,2: 7
6	Past year but not past month, or past 3 years but not past year	47	6,8	1. $R2/(R2+R3)$	1,2: 18 1: 1	1,2: 26	1,2: 2
7	Past 3 years, 30-day frequency missing	7	1,6	1. $R1/(R1+R2+R3)$ 2. $R2/(R1+R2+R3)$ 3. $(R1*D)/(R1+R2+R3)$ 4. $R1*(1-D)*PM/(R1+R2+R3)$	1,2: 0 1: 4	1,2: 0 1: 3	No Cases

¹ The response to CIGDLYMO, the edited response to the “ever daily used” question, technically could be used to subdivide each of the first three missingness patterns into two: one for respondents with CIGDLYMO = 2, and the other for respondents with CIGDLYMO ≠ 2. This was not done, because the benefit derived from this change would likely be insignificant. Respondents with CIGDLYMO = 2 technically have zero probability of being a daily user, so the predictive mean vectors could be simplified by setting D = 0. For example, the predictive mean vector for respondents in Missingness Pattern 2 with CIGDLYMO = 2 might look like this: (1) R1; (2) R2; (3) R3; (4) R1*PM.

² The predictive mean vector components are defined by the following:

1. R1 = P(Past month cigarette use | lifetime cigarette use)
2. R2 = P(Past year but not past month cigarette use | lifetime cigarette use)
3. R3 = P(Past 3 years but not past year cigarette use | lifetime cigarette use)
4. R4 = P(Lifetime but not past 3 years cigarette use | lifetime cigarette use)
5. D = P(Daily cigarette use | past month cigarette use)
6. PM = P(Cigarette use on a given day in the past month | past month cigarette use)

Table E.28 Logical Constraints for Cigar and Smokeless Tobacco Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used within the past 3 years (a recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month user (recency $\neq$ 1)
LogC3	Donor must be a past year user (recency = 1 or 2)
LogC4	Donor must be a past month user (recency = 1)
LogC5	Donor cannot be a past month or past year user (recency $\neq$ 1 or 2)
LogC6	If recipient's age at first use equals recipient's current age, then donor's 30-day frequency (1) cannot be greater than the number of days between recipient's interview date and recipient's date of first use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and recipient's birthday (inclusive)
LogC7	Donor must be a past year but not past month user or a past 3 years but not past year user (recency = 2 or 3)

Table E.29 Likeness Constraints for Cigar and Smokeless Tobacco Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.30 Constraints and Portion of the Predictive Mean Vector for Cigar Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	4	3	1. $R1/(R1+R2)$ 2. $R1*PM/(R1+R2)$	1,2: 0 1: 1	1,2: 3	No Cases
2	Recency and 30-day frequency missing	9	None	1. R1 2. R2 3. R3 4. $R1*PM$	No Cases	1,2: 0 1: 5	1,2: 0 1: 4
3	Past month, 30-day frequency missing	9	4,6	1. PM	1,2: 1 1: 2	1,2: 3	1,2: 3
4	Not past year	67	5	1. $R3/(R3+R4)$	1,2: 9 1: 3	1,2: 36	1,2: 19
5	Not past month	38	2	1. $R2/(R2+R3+R4)$ 2. $R3/(R2+R3+R4)$	1,2: 7 1: 1	1,2: 21	1,2: 7 1: 2
6	Past year but not past month, or past 3 years but not past year	48	7	1. $R2/(R2+R3)$	1,2: 13	1,2: 24	1,2: 11
7	Past 3 years, 30-day frequency missing	1	1	1. $R1/(R1+R2+R3)$ 2. $R2/(R1+R2+R3)$ 3. $(R1*PM)/(R1+R2+R3)$	1,2: 0 1: 1	No Cases	No Cases

¹ The predictive mean vector components are defined by the following:

1. $R1 = P(\text{Past month cigar use} \mid \text{lifetime cigar use})$
2. $R2 = P(\text{Past year but not past month cigar use} \mid \text{lifetime cigar use})$
3. $R3 = P(\text{Past 3 years but not past year cigar use} \mid \text{lifetime cigar use})$
4. $R4 = P(\text{Lifetime but not past 3 years cigar use} \mid \text{lifetime cigar use})$
5. $PM = P(\text{Cigar use on a given day in the past month} \mid \text{past month cigar use})$

Table E.31 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	1	3	1. $R1/(R1+R2)$ 2. $R1*PM/(R1+R2)$	No Cases	1,2: 0 1: 1	No Cases
2	Recency and 30-day frequency missing	4	None	1. R1 2. R2 3. R3 4. $R1*PM$	1,2: 0 1: 1	1,2: 0 1: 2	1,2: 0 1: 1
3	Past month, 30-day frequency missing	1	4,6	1. PM	1,2: 1	No Cases	No Cases
4	Not past year	26	5	1. $R3/(R3+R4)$	1,2: 4 1: 5	1,2: 13	1,2: 4
5	Not past month	14	2	1. $R2/(R2+R3+R4)$ 2. $R3/(R2+R3+R4)$	1,2: 1 1: 6	1,2: 3 1: 2	1,2: 2
6	Past year but not past month, or past 3 years but not past year	17	7	1. $R2/(R2+R3)$	1,2: 4 1: 1	1,2: 12	No Cases
7	Past 3 years, 30-day frequency missing	1	1	1. $R1/(R1+R2+R3)$ 2. $R2/(R1+R2+R3)$ 3. $(R1*PM)/(R1+R2+R3)$	No Cases	1,2: 0 1: 1	No Cases

¹ The predictive mean vector components are defined by the following:

1. $R1 = P(\text{Past month smokeless tobacco use} \mid \text{lifetime smokeless tobacco use})$
2. $R2 = P(\text{Past year but not past month smokeless tobacco use} \mid \text{lifetime smokeless tobacco use})$
3. $R3 = P(\text{Past 3 years but not past year smokeless tobacco use} \mid \text{lifetime smokeless tobacco use})$
4. $R4 = P(\text{Lifetime but not past 3 years smokeless tobacco use} \mid \text{lifetime smokeless tobacco use})$
5. $PM = P(\text{Smokeless tobacco use on a given day in the past month} \mid \text{past month smokeless tobacco use})$

Table E.32 Likeness Constraints for Pipe Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.33 Constraints and Portion of the Predictive Mean Vector for Pipe Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	6	None	1. R1	1,2: 0 1: 1	1,2: 2	1,2: 3

¹ The predictive mean vector components are defined by the following:

1. R1 = P(Past month pipe use | lifetime pipe use)

Table E.34 Logical Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be less than or equal to recipient's maximum possible past year substance frequency of use Recipient's maximum possible substance frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used substance, as determined by the month of first substance use (2) if the maximum period recipient could have used substance is greater than 30, but recipient is a past month substance user with a nonmissing 30-day substance frequency, then the past year substance frequency must be less than or equal to the maximum period (minus the number of days recipient did not use substance in the past month) (3) if recipient is not a past month substance user, then the past year substance frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than or equal to recipient's minimum possible past year substance frequency of use Recipient's minimum possible substance frequency of use in the past year is limited by the following factors: (1) if recipient is a past month substance user, it must be at least as much as the 30-day substance frequency (2) if recipient is not a past month substance user but is a past year substance user, it must be at least 1
LogC3	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than or equal to recipient's 30-day substance use
LogC4	Donor's 30-day substance use must be less than or equal to recipient's maximum number of days could have used substance in past 30 days
LogC5	Donor's 30-day substance use must be greater than or equal to recipient's minimum number of days could have used substance in past 30 days

Table E.34 Logical Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC6	Donor's 30-day substance use must be greater than or equal to recipient's ALCBNG30D if nonmissing
LogC7	Donor's 30-day substance use must be greater than or equal to donor's proportion of past year substance use * recipient's maximum number of days could have used substance in past year [minus 335]
LogC8	Donor must be a past month substance user (substance recency = 1)
LogC9	If recipient's age at first substance use equals recipient's current age, then donor's 30-day substance frequency (1) cannot be greater than recipient's days between recipient's interview date and date of first substance use (inclusive) and (2) cannot be greater than recipient's days between recipient's interview date and birthday (inclusive)
LogC10	If recipient's age at first substance use equals recipient's current age, then (1) donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year cannot be greater than recipient's days between recipient's interview date and date of first substance use (inclusive) and (2) donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year cannot be greater than recipient's days between recipient's interview date and birthday (inclusive)
LogC11	Donor does not have an estimated 30-day frequency
LogC12	If recipient's age at first substance use equals recipient's current age, then (1) donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year cannot be greater than recipient's days between recipient's interview date and date of first substance use (minus 30) and (2) donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year cannot be greater than recipient's days between the interview date and birthday (minus 30)
LogC13	Donor must be a past year but not past month substance user (substance recency = 2)
LogC14	Donor's 30-day binge alcohol frequency (ALCBNG30D) value must be less than or equal to recipient's 30-day substance frequency
LogC15	Donor must have a 30-day binge alcohol frequency (ALCBNG30D) greater than zero if recipient is (1) female and reports usually having four or more drinks on each day she drank during the past 30 days or (2) male and reports usually having five or more drinks on each day he drank during the past 30 days
LogC16	Donor must be a past month or past year but not past month substance user (substance recency = 1 or 2)
LogC17	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than or equal to donor's 30-day substance frequency
LogC18	If recipient's month/year of first substance use data indicate that recipient must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than recipient's 30-day substance frequency
LogC19	If recipient's month/year of first substance use data indicate that recipient must have used substance at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month substance user, then donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than donor's 30-day substance frequency
LogC20	Donor cannot be a past month user (recency ≠ 1)

Table E.34 Logical Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC21	If recipient is a past month user of marijuana and reported past year use of marijuana in a way that recipient did not report using in the past month, then donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than recipient's 30-day substance frequency
LogC22	If recipient is a past month user of marijuana and reported past year use of marijuana in a way that recipient did not report using in the past month, then donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than donor's 30-day substance frequency

Table E.35 Likeness Constraints for Alcohol, Marijuana, and Heroin Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's substance recency must match recipient's substance recency (when nonmissing) ¹

¹ Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.36 Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency

#	Missingness Pattern			Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.	30-Day Freq.				12-17	18-25	26+
1	(Past month)	Missing	Missing	26	1,2,4-11,17,19	1. PM 2. PY	1-3: 2 1,3: 1	1-3: 8 1,3: 2	1-3: 13
2	(Past month)	-	Missing	195	4-6,8,9,11	1. PM	1-3: 7 1,3: 5	1-3: 68 1,3: 2	1-3: 113
3	(Past month)	Missing	-	134	1-3,8,10,18	1. PY	1-3: 22	1-3: 39	1-3: 73
4	(Past year but not past month)	Missing	-	104	1,2,12,13	1. PY	1-3: 43	1-3: 26	1-3: 35
5	Past year	-	Missing	328	4,5,9,11,16	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1-3: 44 1,3: 12	1-3: 102 1,3: 3	1-3: 165 1,3: 2
6	Past year	Missing	Missing	73	1,2,4,5,7,9-12,16,17,19	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-3: 2 1,3: 10	1-3: 8 1,3: 14	1-3: 25 1,3: 14
7	Lifetime (known)	Missing	Missing	252	1,2,4,5,7,9-12,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 50 1,3: 84	1-3: 41 1,3: 53	1-3: 6 1,3: 18

¹ The predictive mean vector components are defined by the following:

1. R1 = P(Past month alcohol use | lifetime alcohol use)
2. R2 = P(Past year but not past month alcohol use | lifetime alcohol use)
3. PM = P(Alcohol use on a given day in the past month | past month alcohol use)
4. PY = P(Alcohol use on a given day in the past year | past year alcohol use)

Table E.37 Constraints and Portion of the Predictive Mean Vector for 30-Day Binge Alcohol Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group					
					Males			Females		
					12-17	18-25	26+	12-17	18-25	26+
1	Missing	787	14,15	1. PMB	1,2: 36 1: 2	1,2: 125	1,2: 184	1,2: 60 1: 3	1,2: 136	1,2: 240 1: 1

¹ The predictive mean vector components are defined by the following:

1. PMB = P(Past month binge alcohol use | past month alcohol use)

Table E.38 Constraints and Portion of the Predictive Mean Vector for Marijuana Recency and Frequency

							Likeness Constraints: Number of Cases, by Age Group		
#	Recency	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	12-17	18-25	26+
		12-Month Freq.	30-Day Freq.						
1	(Past month)	Missing	Missing	15	1,2,4,5,7-11, 17,19,22	1. PM 2. PY	1-3: 2 1,3: 3	1-3: 2 1,3: 1	1-3: 5 1,3: 2
2	(Past month)	-	Missing	8	4,5,8,9,11	1. PM	1-3: 3 1,3: 1	1-3: 3	1-3: 1
3	(Past month)	Missing	-	79	1,2,3,8,10,18, 21	1. PY	1-3: 27 1,3: 1	1-3: 23	1-3: 28
4	(Past year but not past month)	Missing	-	40	1,2,12,13	1. PY	1-3: 18	1-3: 12	1-3: 10
5	Not past month	Missing	-	0	1,2,12,20	1. $R2/(R2+R3)$ 2. $R2*PY/(R2+R3)$	No Cases	No Cases	No Cases
6	Past year	-	Missing	6	4,5,9,11,16	1. $R1/(R1+R2)$ 2. $R1*PM/(R1+R2)$	1-3: 4	1-3: 2	No Cases
7	Past year	Missing	Missing	127	1,2,4,5,7, 9-12,16,17,19	1. $R1/(R1+R2)$ 2. $R1*PM/(R1+R2)$ 3. PY	1-3: 2 1,3: 11	1-3: 14 1,3: 41	1-3: 19 1,3: 40
8	Missing	Missing	Missing	157	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. $R1*PM$ 4. $(R1+R2)*PY$	1-3: 10 1,3: 56	1-3: 6 1,3: 53	1-3: 2 1,3: 30

¹ The predictive mean vector components are defined by the following:

1. $R1 = P(\text{Past month marijuana use} \mid \text{lifetime marijuana use})$
2. $R2 = P(\text{Past year but not past month marijuana use} \mid \text{lifetime marijuana use})$
3. $R3 = P(\text{Lifetime not past year marijuana use} \mid \text{lifetime marijuana use})$
4. $PM = P(\text{Marijuana use on a given day in the past month} \mid \text{past month marijuana use})$
5. $PY = P(\text{Marijuana use on a given day in the past year} \mid \text{past year marijuana use})$

Table E.39 Logical Constraints for Ways of Marijuana Use

Constraint #	Logical Constraint
LogC1	If recipient reported “no” to all nonmissing ways of marijuana use, then donor must have used marijuana in at least one of the missing ways of use

Table E.40 Likeness Constraints for Ways of Marijuana Use

Constraint #	Logical Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor’s predicted means must be within 5 percent of recipient’s predicted means
LikC3	Donor’s value must match recipient’s value for each nonmissing way of marijuana use

Table E.41 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Year

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
1	Missing	-	-	-	-	0	1	1. PS	No Cases	No Cases	No Cases
2	-	Missing	-	-	-	0	1	1. PV	No Cases	No Cases	No Cases
3	Missing	Missing	-	-	-	0	1	1. PS 2. PV	No Cases	No Cases	No Cases
4	-	-	Missing	-	-	0	1	1. PD	No Cases	No Cases	No Cases
5	Missing	-	Missing	-	-	0	1	1. PS 2. PD	No Cases	No Cases	No Cases
6	-	Missing	Missing	-	-	0	1	1. PV 2. PD	No Cases	No Cases	No Cases
7	Missing	Missing	Missing	-	-	0	1	1. PS 2. PV 3. PD	No Cases	No Cases	No Cases
8	-	-	-	Missing	-	0	1	1. PE	No Cases	No Cases	No Cases
9	Missing	-	-	Missing	-	0	1	1. PS 2. PE	No Cases	No Cases	No Cases
10	-	Missing	-	Missing	-	0	1	1. PV 2. PE	No Cases	No Cases	No Cases
11	Missing	Missing	-	Missing	-	0	1	1. PS 2. PV 3. PE	No Cases	No Cases	No Cases
12	-	-	Missing	Missing	-	0	1	1. PD 2. PE	No Cases	No Cases	No Cases
13	Missing	-	Missing	Missing	-	0	1	1. PS 2. PD 3. PE	No Cases	No Cases	No Cases
14	-	Missing	Missing	Missing	-	0	1	1. PV 2. PD 3. PE	No Cases	No Cases	No Cases
15	Missing	Missing	Missing	Missing	-	0	1	1. PS 2. PV 3. PD 4. PE	No Cases	No Cases	No Cases
16	-	-	-	-	Missing	0	1	1. PO	No Cases	No Cases	No Cases

**Table E.41 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Year
(continued)**

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
17	Missing	-	-	-	Missing	0	1	1. PS 2. PO	No Cases	No Cases	No Cases
18	-	Missing	-	-	Missing	0	1	1. PV 2. PO	No Cases	No Cases	No Cases
19	Missing	Missing	-	-	Missing	0	1	1. PS 2. PV 3. PO	No Cases	No Cases	No Cases
20	-	-	Missing	-	Missing	1	1	1. PD 2. PO	No Cases	No Cases	1-3: 0 2,3: 0 1,2: 1
21	Missing	-	Missing	-	Missing	0	1	1. PS 2. PD 3. PO	No Cases	No Cases	No Cases
22	-	Missing	Missing	-	Missing	0	1	1. PV 2. PD 3. PO	No Cases	No Cases	No Cases
23	Missing	Missing	Missing	-	Missing	0	1	1. PS 2. PV 3. PD 4. PO	No Cases	No Cases	No Cases
24	-	-	-	Missing	Missing	0	1	1. PE 2. PO	No Cases	No Cases	No Cases
25	Missing	-	-	Missing	Missing	0	1	1. PS 2. PE 3. PO	No Cases	No Cases	No Cases
26	-	Missing	-	Missing	Missing	1	1	1. PV 2. PE 3. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	No Cases	No Cases

**Table E.41 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Year
(continued)**

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
27	Missing	Missing	-	Missing	Missing	0	1	1. PS 2. PV 3. PE 4. PO	No Cases	No Cases	No Cases
28	-	-	Missing	Missing	Missing	0	1	1. PD 2. PE 3. PO	No Cases	No Cases	No Cases
29	Missing	-	Missing	Missing	Missing	3	1	1. PS 2. PD 3. PE 4. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 2	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	No Cases
30	-	Missing	Missing	Missing	Missing	4	1	1. PV 2. PD 3. PE 4. PO	No Cases	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 3
31	Missing	Missing	Missing	Missing	Missing	147	1	1. PS 2. PV 3. PD 4. PE 5. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 68	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 53	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 26

¹ The predictive mean vector components are defined by the following:

- 1. PS = P(Smoked marijuana in past year | past year marijuana use)
- 2. PV = P(Vaped marijuana in past year | past year marijuana use)
- 3. PD = P(Dabbed marijuana in past year | past year marijuana use)
- 4. PE = P(Ate marijuana in past year | past year marijuana use)
- 5. PO = P(Used marijuana in some other way in past year | past year marijuana use)

Table E.42 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Month

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
1	Missing	-	-	-	-	0	1	1. PS	No Cases	No Cases	No Cases
2	-	Missing	-	-	-	0	1	1. PV	No Cases	No Cases	No Cases
3	Missing	Missing	-	-	-	0	1	1. PS 2. PV	No Cases	No Cases	No Cases
4	-	-	Missing	-	-	0	1	1. PD	No Cases	No Cases	No Cases
5	Missing	-	Missing	-	-	0	1	1. PS 2. PD	No Cases	No Cases	No Cases
6	-	Missing	Missing	-	-	0	1	1. PV 2. PD	No Cases	No Cases	No Cases
7	Missing	Missing	Missing	-	-	0	1	1. PS 2. PV 3. PD	No Cases	No Cases	No Cases
8	-	-	-	Missing	-	0	1	1. PE	No Cases	No Cases	No Cases
9	Missing	-	-	Missing	-	0	1	1. PS 2. PE	No Cases	No Cases	No Cases
10	-	Missing	-	Missing	-	0	1	1. PV 2. PE	No Cases	No Cases	No Cases
11	Missing	Missing	-	Missing	-	0	1	1. PS 2. PV 3. PE	No Cases	No Cases	No Cases
12	-	-	Missing	Missing	-	0	1	1. PD 2. PE	No Cases	No Cases	No Cases
13	Missing	-	Missing	Missing	-	0	1	1. PS 2. PD 3. PE	No Cases	No Cases	No Cases
14	-	Missing	Missing	Missing	-	0	1	1. PV 2. PD 3. PE	No Cases	No Cases	No Cases
15	Missing	Missing	Missing	Missing	-	0	1	1. PS 2. PV 3. PD 4. PE	No Cases	No Cases	No Cases
16	-	-	-	-	Missing	0	1	1. PO	No Cases	No Cases	No Cases

Table E.42 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Month
(continued)

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
17	Missing	-	-	-	Missing	0	1	1. PS 2. PO	No Cases	No Cases	No Cases
18	-	Missing	-	-	Missing	0	1	1. PV 2. PO	No Cases	No Cases	No Cases
19	Missing	Missing	-	-	Missing	0	1	1. PS 2. PV 3. PO	No Cases	No Cases	No Cases
20	-	-	Missing	-	Missing	1	1	1. PD 2. PO	No Cases	No Cases	1-3: 0 2,3: 0 1,2: 1
21	Missing	-	Missing	-	Missing	0	1	1. PS 2. PD 3. PO	No Cases	No Cases	No Cases
22	-	Missing	Missing	-	Missing	0	1	1. PV 2. PD 3. PO	No Cases	No Cases	No Cases
23	Missing	Missing	Missing	-	Missing	0	1	1. PS 2. PV 3. PD 4. PO	No Cases	No Cases	No Cases
24	-	-	-	Missing	Missing	0	1	1. PE 2. PO	No Cases	No Cases	No Cases
25	Missing	-	-	Missing	Missing	0	1	1. PS 2. PE 3. PO	No Cases	No Cases	No Cases
26	-	Missing	-	Missing	Missing	1	1	1. PV 2. PE 3. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	No Cases	No Cases

Table E.42 Constraints and Portion of the Predictive Mean Vector for Ways of Marijuana Use in Past Month (continued)

#	Missingness Pattern					Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Smoking	Vaping	Dabbing	Eating	Other				12-17	18-25	26+
27	Missing	Missing	-	Missing	Missing	0	1	1. PS 2. PV 3. PE 4. PO	No Cases	No Cases	No Cases
28	-	-	Missing	Missing	Missing	0	1	1. PD 2. PE 3. PO	No Cases	No Cases	No Cases
29	Missing	-	Missing	Missing	Missing	3	1	1. PS 2. PD 3. PE 4. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 2	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	No Cases
30	-	Missing	Missing	Missing	Missing	4	1	1. PV 2. PD 3. PE 4. PO	No Cases	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 1	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 3
31	Missing	Missing	Missing	Missing	Missing	147	1	1. PS 2. PV 3. PD 4. PE 5. PO	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 68	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 53	1-3: 0 2,3: 0 1,2: 0 2: 0 1,3: 26

¹ The predictive mean vector components are defined by the following:
1. PS = P(Smoked marijuana in past month | past month marijuana use)
2. PV = P(Vaped marijuana in past month | past month marijuana use)
3. PD = P(Dabbed marijuana in past month | past month marijuana use)
4. PE = P(Ate marijuana in past month | past month marijuana use)
5. PO = P(Used marijuana in some other way in past month | past month marijuana use)

Table E.43 Constraints and Portion of the Predictive Mean Vector for Heroin Recency and Frequency

							Likeness Constraints: Number of Cases, by Age Group		
#	Recency	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	12-17	18-25	26+
		12-Month Freq.	30-Day Freq.						
1	(Past month)	Missing	Missing	0	1,2,4,5,7-11, 17,19	1. PM 2. PY	No Cases	No Cases	No Cases
2	(Past month)	-	Missing	1	4,5,8,9,11	1. PM	No Cases	No Cases	1-3: 0 1,3: 1
3	(Past month)	Missing	-	1	1-3,8,10,18	1. PY	No Cases	No Cases	1-3: 1
4	(Past year but not past month)	Missing	-	0	1,2,12,13	1. PY	No Cases	No Cases	No Cases
5	Past year	-	Missing	1	4,5,9,11,16	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-3: 0 1,3: 1	No Cases
6	Past year	Missing	Missing	6	1,2,4,5,7,9-12, 16,17,19	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	1-3: 0 1,3: 6
7	Missing	Missing	Missing	7	1,2,4,5,7,9- 12,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 2	1-3: 0 1,3: 1	1-3: 0 1,3: 4

¹ The predictive mean vector components are defined by the following:

1. R1 = P(Past month heroin use | lifetime heroin use)
2. R2 = P(Past year but not past month heroin use | lifetime heroin use)
3. PM = P(Heroin use on a given day in the past month | past month heroin use)
4. PY = P(Heroin use on a given day in the past year | past year heroin use)

Table E.44 Logical Constraints for Cocaine Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must be a past month cocaine user (cocaine recency = 1)
LogC2	Donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to recipient's maximum possible past year cocaine frequency of use Recipient's maximum possible cocaine frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used cocaine, as determined by the month of first cocaine use (2) if the maximum period recipient could have used cocaine is greater than 30, but recipient is a past month cocaine user with a nonmissing 30-day cocaine frequency, then the past year cocaine frequency must be less than or equal to the maximum period (the number of days recipient did not use cocaine in the past month) (3) if recipient is not a past month cocaine user, then the past year cocaine frequency must be less than or equal to the maximum period (minus 30)
LogC3	Donor's proportion of past year cocaine use * recipient's minimum number of days could have used cocaine in the past year must be greater than or equal to recipient's minimum possible past year cocaine frequency of use Recipient's minimum possible cocaine frequency of use in the past year is limited by the following factors: (1) if recipient is a past month cocaine user, it must be at least as much as the 30-day cocaine frequency (2) if recipient is not a past month cocaine user but is a past year cocaine user, it must be at least 1 (3) if recipient has a valid value for the 12-month crack frequency, it must be at least as much as the recipient's 12-month crack frequency
LogC4	If recipient's age at first cocaine use equals recipient's current age, and (1) donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first substance use (inclusive) (3) donor is a past year but not past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30) (4) donor is a past year but not past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first substance use (minus 30)
LogC5	If recipient's age at first cocaine use equals recipient's current age, then (1) donor's 30-day cocaine use must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) donor's 30-day cocaine use must be less than or equal to the number of days between recipient's interview date and date of first substance use (inclusive)
LogC6	Donor's 30-day cocaine use must be less than or equal to recipient's maximum number of days could have used cocaine in past 30 days
LogC7	Donor's 30-day cocaine use must be greater than or equal to recipient's minimum number of days could have used cocaine in past 30 days

Table E.44 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC8	Donor's 30-day cocaine use must be greater than or equal to donor's proportion of past year substance use * recipient's maximum number of days could have used substance in past year (minus 335)
LogC9	Donor does not have an estimated 30-day frequency for cocaine
LogC10	Donor must be a past month or past year but not past month cocaine user (cocaine recency = 1 or 2)
LogC11	Donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be less than or equal to recipient's maximum possible past year crack frequency of use Recipient's maximum possible crack frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used crack, as determined by the month of first crack use (2) if the maximum period recipient could have used crack is greater than 30, but recipient is a past month crack user with a nonmissing 30-day crack frequency, then the past year crack frequency must be less than or equal to the maximum period (the number of days recipient did not use in the past month) (3) if recipient is not a past month crack user, then the past year crack frequency must be less than or equal to the maximum period (minus 30) (4) if recipient has a valid value for the 12-month cocaine frequency, it must be less than or equal to the recipient's 12-month cocaine frequency
LogC12	Donor's proportion of past year crack use * recipient's minimum number of days could have used crack in the past year must be greater than or equal to recipient's minimum possible past year crack frequency of use Recipient's minimum possible crack frequency of use in the past year is limited by the following factors: (1) if recipient is a past month crack user, it must be at least as much as the 30-day crack frequency (2) if recipient is not a past month crack user but is a past year crack user, it must be at least 1
LogC13	If recipient's age at first crack use equals recipient's current age, and (1) donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first substance use (inclusive) (3) donor is a past year but not past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30) (4) donor is a past year but not past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and date of first substance use (minus 30)
LogC14	Donor's 30-day crack use must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC15	Donor's 30-day crack use must be less than or equal to recipient's maximum number of days could have used crack in past 30 days

Table E.44 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC16	Donor's 30-day crack use must be greater than or equal to donor's proportion of past year substance use * recipient's maximum number of days could have used substance in past year (minus 335)
LogC17	Donor's 30-day crack use must be greater than or equal to recipient's minimum number of days could have used crack in past 30 days
LogC18	Donor does not have an estimated 30-day frequency for crack
LogC19	Donor must be a past month crack user (crack recency = 1)
LogC20	Donor must be a past month or past year not past month crack user (crack recency = 1, 2)
LogC21	Donor must be a crack user (crack recency ≠ 91)
LogC22	If recipient's month/year of first use data for cocaine indicate that recipient must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), and recipient is a past month user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than recipient's 30-day cocaine frequency
LogC23	If recipient's month/year of first use data for cocaine indicate that recipient must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), and donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than donor's 30-day cocaine frequency
LogC24	If recipient's month/year of first use data for crack indicate that recipient must have used crack at least once in the interval (1 year before interview, 30 days before interview), and recipient is a past month user, then donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be greater than recipient's 30-day crack frequency
LogC25	If recipient's month/year of first use data for crack indicate that recipient must have used crack at least once in the interval (1 year before interview, 30 days before interview), and donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used in the past year must be greater than donor's 30-day crack frequency
LogC26	Donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be less than or equal to donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year
LogC27	If recipient is a past month cocaine user or cocaine recency is missing, donor is a past month cocaine user, and donor is a past year not past month crack user, then donor's proportion of days used crack in the last year * recipient's maximum period used crack in the past year cannot be larger than recipient's 12-month frequency minus donor's 30-day frequency

Table E.45 Likeness Constraints for Cocaine Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's cocaine recency must match recipient's substance recency (when nonmissing) ¹
LikC4	Donor's crack recency agrees with recipient's crack recency (when nonmissing)

¹ Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.46 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency

#	Missingness Pattern							Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Total Number of Cases			12-17	18-25	26+
1	(Past year)	-	Missing	-	-	-	20	2-4,10,22	1. PY	1-4: 1	1-4: 1 1-3: 2 1,3: 1	1-4: 2 1-3: 5 1,3: 8
2	(Past year)	(Past year)	-	Missing	-	-	0	11-13,20,24	1. PY	No Cases	No Cases	No Cases
3	(Past year)	(Past year)	Missing	Missing	-	-	2	2-4,10-13,20,22,24,26	1. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 1
4	(Past month)	-	-	-	Missing	-	7	1,5-7,9	1. PM	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 1 1-3: 3 1,3: 2
5	(Past month)	(Past month)	-	-	-	Missing	0	14-16,18,19	1. PM	No Cases	No Cases	No Cases
6	(Past month)	(Past month)	-	-	Missing	Missing	0	1,5-7,9,14-16,18,19	1. PM	No Cases	No Cases	No Cases
7	(Past month)	-	Missing	-	Missing	-	4	1-9,23	1. PM 2. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 4
8	(Past month)	(Past year)	-	Missing	Missing	-	0	1,5-7,9,11-13,20,24	1. PM 2. PY	No Cases	No Cases	No Cases
9	(Past month)	(Past year)	Missing	Missing	Missing	-	0	1-9,11-13,20,23,24,26	1. PM 2. PY	No Cases	No Cases	No Cases
10	(Past month)	(Past month)	Missing	-	-	Missing	0	2-4,14-16,18,19,22	1. PM 2. PY	No Cases	No Cases	No Cases
11	(Past month)	(Past month)	-	Missing	-	Missing	0	11-19,25	1. PM 2. PY	No Cases	No Cases	No Cases
12	(Past month)	(Past month)	Missing	Missing	-	Missing	0	2-4,11-19,22,25,26	1. PM 2. PY	No Cases	No Cases	No Cases

Table E.46 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern							Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.	Total Number of Cases			12-17	18-25	26+
13	(Past month)	(Past month)	Missing		Missing	Missing	0	1-9,14-16,18,19,23	1. PM 2. PY	No Cases	No Cases	No Cases
14	(Past month)	(Past month)	-	Missing	Missing	Missing	0	1,5-7,9,11-19,25	1. PM 2. PY	No Cases	No Cases	No Cases
15	(Past month)	(Past month)	Missing	Missing	Missing	Missing	1	1-9,11-19,23,25,26	1. PM 2. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
16	(Past month)	Past year	-	-	-	Missing	0	14-16,18,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	No Cases	No Cases	No Cases
17	(Past month)	Past year	-	-	Missing	Missing	0	1,5-7,9,14-16,18,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	No Cases	No Cases	No Cases
18	(Past month)	Past year	-	Missing	-	Missing	0	11-18,20,25	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
19	(Past month)	Past year	-	Missing	Missing	Missing	0	1,5-7,9,11-18,20,25,27	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
20	(Past month)	Missing	-	Missing	-	Missing	2	11-18,21	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 1
21	(Past month)	Missing	-	Missing	Missing	Missing	0	1,5-7,9,11-18,21,27	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases

Table E.46 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12- Mo. Freq.	Cocaine 30-Day Freq.	Crack 30- Day Freq.				12-17	18-25	26+
22	(Past month)	Missing	Missing	Missing	Missing	Missing	0	1-9,11-18,21, 23,26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
23	Missing	-	Missing	-	Missing		25	2-9	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 13	1-4: 0 1-3: 0 1,3: 10
24	Missing	Missing	Missing	Missing	Missing	Missing	7	2-9,11-18,21, 26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 4
25	Past year	-	-	-	Missing	-	13	5-7,9,10	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-4: 2 1-3: 0 1,3: 4	1-4: 3 1-3: 2 1,3: 2
26	Past year	Past year	-	-	Missing	Missing	0	5-7,9,10, 14-16,18,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
27	Past year	-	Missing	-	Missing	-	3	2-10,23	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 2	1-4: 0 1-3: 0 1,3: 1
28	Past year	Past year	Missing	-	Missing	Missing	1	2-10,14-16, 18,20,23	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
29	Past year	Past year	-	Missing	Missing	Missing	2	5-7,9-18,20, 25,27	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 1

Table E.46 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
30	Past year	Past year	Missing	Missing	Missing	Missing	1	2-18,20,23, 25,26	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
31	Past year	Missing	-	Missing	Missing	Missing	0	5-7,9-18,21, 27	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
32	Past year	Missing	Missing	Missing	Missing	Missing	0	2-18,21,23, 26	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases
33	(Past month)	Missing	Missing	Missing	-	Missing	0	2-4,10-18,21, 22,26	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
34	Past year	(Past year but not past month)	Missing	Missing	Missing	-	0	2-13,20,23, 26	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	No Cases	No Cases	No Cases

NOTE: Cocaine users included crack users and cocaine users who were not crack users.

¹ The predictive mean vector components are defined by the following:

1. R1 = P(Past month cocaine use | lifetime cocaine use)
2. R2 = P(Past year but not past month cocaine use | lifetime cocaine use)
3. PM = P(Cocaine use on a given day in the past month | past month use of cocaine)
4. PY = P(Cocaine use on a given day in the past year | past year use of cocaine)

E.3.1.3 Age at First Use**Table E.47 Logical Constraints for Age at First Use, Univariate Assignment**

Constraint #	Logical Constraint
LogC1	Donor's age at first substance use must be less than or equal to recipient's age
LogC2	If recipient was not a past year substance user, then donor cannot have age at first substance use equal to recipient's age
LogC3	If recipient was not a past 3 years substance user, then donor cannot have age at first substance use equal to recipient's age, recipient's age minus 1, or recipient's age minus 2
LogC4	If recipient's 12-month substance frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first substance use equal to recipient's age
LogC5	If recipient's 30-day substance frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first substance use equal to recipient's age
LogC6	If recipient's most recent birthday was within the past 30 days, and recipient's substance recency was past year but not past month, then donor cannot have age at first substance use equal to recipient's age
LogC7	If recipient's substance recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month substance frequency was less than 30, then donor cannot have age at first substance use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first substance use must be earlier than recipient's most recent birthday.)
LogC8	Donor's age at first cigarette use must be less than or equal to recipient's age at first daily cigarette use (if existing)
LogC9	Donor's age at first daily cigarette use must be greater than or equal to recipient's imputation-revised age at first cigarette use
LogC10	If recipient's most recent birthday was within the past 30 days, and recipient used substance in a way in the past year that recipient did not use in the past month, then donor cannot have age at first substance use equal to recipient's age

Table E.48 Likeness Constraints for Age at First Use, Univariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient was a past year substance user, then donor was a past year substance user; if recipient was a past 3 years but not past year substance user, then donor was the same; if recipient was a lifetime but not past 3 years substance user, then donor was the same
LikC4	Age of donor = age of recipient
LikC5	Age of donor must be greater than or equal to age of recipient
LikC6	If recipient was a past year substance user, then donor was a past year substance user; if recipient was not a past year substance user, then donor was the same
LikC7	If recipient was a past 3 years but not past year substance user, then donor was not a past year substance user; if recipient was a lifetime but not past 3 years substance user, then donor was the same
LikC8	If recipient was not a past year substance user, then donor was the same

Table E.49 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	245	1-3,5,6,8	1. PrAFU	1-4: 68 2-4: 3	1-4: 89 2-4: 1	1-4: 81 2-4: 2 3,4: 0 4,7: 0 5,7: 1

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first use

Table E.50 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Daily Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	81	1-3,5,6,9	1. PrCigAFU	1-4: 3 2-4: 3 3,4: 0 4,7: 0 5,7: 1	1-4: 17 2-4: 1 3,4: 1	1-4: 48 2-4: 3 3,4: 4

¹ The predictive mean vector components are defined by the following:
1. PrCigAFU = Predicted cigarette age at first daily use

Table E.51 Constraints and Portion of the Predictive Mean Vector for Cigar Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	260	1-3,5,6	1. PrAFU	1-4: 24 2-4: 1 3,4: 3	1-4: 80 2-4: 1 3,4: 1	1-4: 139 2-4: 7 3,4: 4

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigar age at first use

Table E.52 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	97	1-3,5,6	1. PrAFU	1-4: 10 2-4: 8 3,4: 5	1-4: 27 2-4: 1	1-4: 39 2-4: 6 3,4: 0 4,7: 1

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted smokeless tobacco age at first use

Table E.53 Constraints and Portion of the Predictive Mean Vector for Alcohol Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	464	1,2,4,6,7	1. PrAFU	1,2,4,6: 153	1,2,4,6: 107	1,2,4,6: 195 2,4,6: 6 4,6: 3

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted alcohol age at first use

Table E.54 Constraints and Portion of the Predictive Mean Vector for Marijuana Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	198	1,2,4,6,7,10	1. PrAFU	1,2,4,6: 62	1,2,4,6: 55	1,2,4,6: 78 2,4,6: 1 4,6: 2

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted marijuana age at first use

Table E.55 Constraints and Portion of the Predictive Mean Vector for Heroin Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	11	1,2,4,6,7	1. PrAFU	1,2,4,6: 0 2,4,6: 0 4,6: 1	1,2,4,6: 2	1,2,4,6: 1 2,4,6: 4 4,6: 3

¹ The predictive mean vector components are defined by the following:

1. PrAFU = Predicted heroin age at first use

Table E.56 Logical Constraints for Age at First Use, Multivariate Assignment

Constraint #	Logical Constraint
LogC1	For each parent or child drug for which recipient was missing age at first use, donor's age at first use must be less than or equal to recipient's age
LogC2	For each parent or child drug for which recipient was missing age at first use, if recipient was not a past year user, then donor cannot have age at first use equal to recipient's age
LogC3	For each child drug for which recipient was missing age at first use, donor must be a lifetime user
LogC4	If recipient was missing the age at first use for the parent drug, and recipient's parent drug 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have parent drug age at first use equal to recipient's age
LogC5	For each child drug for which recipient was missing age at first use, if recipient's 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first use equal to recipient's age
LogC6	If recipient was missing the age at first use for the parent drug, then donor must have parent drug age at first use less than or equal to the recipient's child drug(s) age(s) at first use (if existing)
LogC7	For each parent or child drug for which recipient was missing age at first use, if recipient's most recent birthday was within the past 30 days, and recipient's recency was past year but not past month, then donor cannot have age at first use equal to recipient's age
LogC8	If recipient was missing the age at first use for the parent drug, and recipient's parent drug recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's parent drug 12-month frequency was less than 30, then donor cannot have parent drug age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)
LogC9	For each child drug for which recipient was missing age at first use, if recipient's recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month frequency was less than 30, then donor cannot have age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)
LogC10	For each child drug for which recipient was missing age at first use, donor must have age at first use greater than or equal to recipient's parent drug age at first use (if existing)

Table E.57 Likeness Constraints for Age at First Use, Multivariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = age of recipient
LikC4	Donor agrees with recipient with respect to lifetime use for child drug(s)
LikC5	For the parent drug, if recipient was a past year user, then donor was a past year user
LikC6	For the parent drug, if recipient was a lifetime but not past year user, then donor was a lifetime but not past year user
LikC7	For each child drug, if recipient was a past year user, then donor was a past year user
LikC8	For each child drug, if recipient was a lifetime but not past year user, then donor was not a past year user
LikC9	Donor was at least as old as recipient, but no more than 20 years older than recipient
LikC10	Donor was no more than 20 years older than recipient

Table E.58 Constraints and Portion of the Predictive Mean Vector for Cocaine Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	46	1-10	1. PrCocAFU	1-8: 0 2-8: 1	1-8: 10	1-8: 26 2-8: 4 2,3,5-8: 0 3,5-8: 3 3,6,8: 2

¹ The predictive mean vector components are defined by the following:
1. PrCocAFU = Predicted cocaine age at first use

E.3.2 modPMN

E.3.2.1 Lifetime Substance Use

Table E.59 Imputation Details for Inhalants Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	191	1. LF1 2. LF2
Total cases requiring imputation	191	-

¹ The predictive mean vector components are defined by the following:
1. LF1 = P(Lifetime use of inhalants)
2. LF2 = P(Never used inhalants)

Table E.60 Imputation Details for Methamphetamine Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	95	1. LF1 2. LF2
Total cases requiring imputation	95	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of methamphetamine)
2. LF2 = P(Never used methamphetamine)

Table E.61 Imputation Details for Hallucinogens Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	290	1. LF1 2. LF2
Total cases requiring imputation	290	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of hallucinogens)
2. LF2 = P(Never used hallucinogens)

There were a large number of missingness patterns for lifetime hallucinogen child drug use. The usable case rule (Section 2.2.3) required a certain minimum response pattern, but other than that any combination of the lifetime hallucinogen child drug variables may have been missing for a given respondent record.

The probabilities associated with the seven lifetime hallucinogen child drug variables listed in [Table E.62](#) formed the full predictive mean vector. Only the probabilities associated with the lifetime hallucinogen child drug variables for which values were missing were used in the predictive mean vector for each item nonrespondent. Although lifetime hallucinogen child drug variables were imputed using PMN methodology, they have been included in this section to allow proximity to the parent measure, hallucinogens lifetime.

Table E.62 Elements of Full Predictive Mean Vector for Hallucinogen Child Lifetime Drug Use

Lifetime Drug Use	Predicted Mean
LSD Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
PCP Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
Ecstasy Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
Ketamine Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
DMT/AMT/Foxy Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
<i>Salvia divinorum</i> Lifetime	P(Lifetime User Lifetime User of Hallucinogens)
Other Hallucinogens Lifetime ¹	P(Lifetime User Lifetime User of Hallucinogens)

¹ Other hallucinogens for lifetime directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not previously included in the lifetime predictive mean vector.

Table E.63 Logical Constraints for Hallucinogen Child Lifetime Drug Use

Constraint #	Logical Constraint
LogC1	Donor must be a lifetime user of at least one hallucinogen child drug that the recipient is missing if recipient is a nonuser of all nonmissing drugs or is missing all child drugs

Table E.64 Likeness Constraints for Hallucinogen Child Lifetime Drug Use

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Lifetime use of donor = lifetime use of recipient, for each nonmissing lifetime indicator

Table E.65 Constraints and Portion of the Predictive Mean Vector for Hallucinogen Child Lifetime Drug Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing one or more of the lifetime drug use variables	69	1	1. LF for gate question associated with each missing variable	1-3: 1 1,2: 2 1: 7	1-3: 8 1,2: 0 1: 3	1-3: 36 1,2: 7 1: 5

¹ The predictive mean vector components are defined by the following:

1. LF = P(Lifetime user of x | lifetime user of hallucinogens), where x is the child drug associated with the missing gate question

Table E.66 Imputation Details for Lifetime Vaping of Nicotine or Tobacco

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	19	1. LF1 2. LF2
Total cases requiring imputation	19	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime vaping of nicotine or tobacco)
2. LF2 = P(Never vaped nicotine or tobacco)

Table E.67 Imputation Details for Lifetime Flavoring Vaping

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,696	1. LF1 2. LF2
Total cases requiring imputation	1,696	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime flavoring vaping)
2. LF2 = P(Never flavoring vaped)

Table E.68 Imputation Details for Illegally Made Fentanyl Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,616	1. LF1 2. LF2
Total cases requiring imputation	1,616	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of illegally made fentanyl)
2. LF2 = P(Never used illegally made fentanyl)

Table E.69 Imputation Details for Kratom Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,805	1. LF1 2. LF2
Total cases requiring imputation	1,805	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of kratom)
2. LF2 = P(Never used kratom)

Table E.70 Imputation Details for Synthetic Marijuana Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,650	1. LF1 2. LF2
Total cases requiring imputation	1,650	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of synthetic marijuana)
2. LF2 = P(Never used synthetic marijuana)

Table E.71 Imputation Details for Synthetic Stimulants Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,603	1. LF1 2. LF2
Total cases requiring imputation	1,603	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of synthetic stimulants)
2. LF2 = P(Never used synthetic stimulants)

Table E.72 Imputation Details for CBD or Hemp Lifetime Use

Edited Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	519	1. LF1 2. LF2
Total cases requiring imputation	519	-

¹ The predictive mean vector components are defined by the following:

1. LF1 = P(Lifetime use of CBD or hemp)
2. LF2 = P(Never used CBD or hemp)

E.3.2.2 Recency and Frequency of Substance Use/Misuse

Many tables in this section abbreviate certain phrases. "Recency" is an abbreviation for "recency of use," and "frequency" or "Freq." is an abbreviation for "frequency of use."

An entry of “Missing” within “Missingness Pattern” indicates that all information was missing. Other entries for missingness patterns provide the available information, indicating that this information was partially missing. For example, an entry of “Past year” indicates that the individual used the substance of interest in the past year, but the specific recency of past month versus past year not past month must be imputed.

E.3.2.2.1 Recency of Any Use

Table E.73 Imputation Details for Sedatives Recency of Any Use

Edited Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Lifetime use	22	1. $RU1/(RU1+RU2)$ 2. $RU2/(RU1+RU2)$
Not past year (lifetime use unknown)	198	1. $RU2/(RU2+RU3)$ 2. $RU3/(RU2+RU3)$
Completely missing	311	1. RU1 2. RU2 3. RU3
Total cases requiring imputation	531	-

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of sedatives)
2. RU2 = P(Lifetime use of sedatives)
3. RU3 = P(Never used sedatives)

Table E.74 Imputation Details for Stimulants Recency of Any Use

Edited Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Lifetime use	14	1. $RU1/(RU1+RU2)$ 2. $RU2/(RU1+RU2)$
Not past year (lifetime use unknown)	142	1. $RU2/(RU2+RU3)$ 2. $RU3/(RU2+RU3)$
Completely missing	269	1. RU1 2. RU2 3. RU3
Total cases requiring imputation	425	-

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of stimulants)
2. RU2 = P(Lifetime use of stimulants)
3. RU3 = P(Never used stimulants)

Table E.75 Imputation Details for Tranquilizers Recency of Any Use

Edited Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Lifetime use	24	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Not past year (lifetime use unknown)	293	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
Completely missing	267	1. RU1 2. RU2 3. RU3
Total cases requiring imputation	584	-

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of tranquilizers)
2. RU2 = P(Lifetime use of tranquilizers)
3. RU3 = P(Never used tranquilizers)

Table E.76 Imputation Details for Pain Relievers Recency of Any Use

Edited Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Lifetime use	169	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Not past year (lifetime use unknown)	391	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
Completely missing	237	1. RU1 2. RU2 3. RU3
Total cases requiring imputation	797	-

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of pain relievers)
2. RU2 = P(Lifetime use of pain relievers)
3. RU3 = P(Never used pain relievers)

Table E.77 Imputation Details for OxyContin Recency of Any Use

Edited Recency of Any Use	Restrictions Based on Imputation-Revised Pain Relievers Any Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	212	1. RU1 2. RU2
	No past year any use	115	N/A
Total cases requiring imputation	-	327	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of OxyContin | past year use of pain relievers)
2. RU2 = P(No past year use of OxyContin | past year use of pain relievers)

Table E.78 Imputation Details for Prescription Fentanyl Recency of Any Use

Edited Recency of Any Use	Restrictions Based on Imputation-Revised Pain Relievers Any Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	180	1. RU1 2. RU2
	No past year any use	107	N/A
Total cases requiring imputation	-	287	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of prescription fentanyl | past year use of pain relievers)
2. RU2 = P(No past year use of prescription fentanyl | past year use of pain relievers)

Table E.79 Imputation Details for Benzodiazepines Recency of Any Use

Edited Recency of Any Use	Restrictions Based on Imputation-Revised Sedatives Any Use and Imputation-Revised Tranquilizers Any Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	73	1. RU1 2. RU2
	No past year any use	266	N/A
Total cases requiring imputation	-	339	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past year use of Benzodiazepines | past year use of sedatives or past year use of tranquilizers)
2. RU2 = P(No past year use of Benzodiazepines | past year use of sedatives or past year use of tranquilizers)

E.3.2.2.2 Recency of Use/Misuse**Table E.80 Imputation Details for Inhalants Recency of Use**

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	13	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	178	1. RU1 2. RU2 3. RU3
	Never used	176	N/A
Total cases requiring imputation	-	367	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of inhalants | lifetime use of inhalants)
2. RU2 = P(Past year but not past month use of inhalants | lifetime use of inhalants)
3. RU3 = P(Lifetime but not past year use of inhalants | lifetime use of inhalants)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.81 Imputation Details for Methamphetamine Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	14	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	17	1. RU1 2. RU2 3. RU3
	Never used	90	N/A
Total cases requiring imputation	-	121	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of methamphetamine | lifetime use of methamphetamine)
2. RU2 = P(Past year but not past month use of methamphetamine | lifetime use of methamphetamine)
3. RU3 = P(Lifetime but not past year use of methamphetamine | lifetime use of methamphetamine)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.82 Imputation Details for Hallucinogens Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	39	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	126	1. RU1 2. RU2 3. RU3
	Never used	271	N/A
Total cases requiring imputation	-	436	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of hallucinogens | lifetime use of hallucinogens)
2. RU2 = P(Past year but not past month use of hallucinogens | lifetime use of hallucinogens)
3. RU3 = P(Lifetime but not past year use of hallucinogens | lifetime use of hallucinogens)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.83 Imputation Details for LSD Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	5	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	1	N/A
		Past year, not past month	N/A	1	N/A
Completely missing ³	None	None	None	7	1. RU1 2. RU2 3. RU3
			Past month	1	N/A
		Not past month	None	10	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	2	N/A
		Lifetime, not past year	N/A	30	N/A
		Never used	N/A	54	N/A
Total cases requiring imputation	-	-	-	111	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used LSD during lifetime and LSD recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only LSD recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of LSD | lifetime use of LSD)

2. RU2 = P(Past year but not past month use of LSD | lifetime use of LSD)

3. RU3 = P(Lifetime but not past year use of LSD | lifetime use of LSD)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.84 Imputation Details for PCP Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	0	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	0	N/A
		Past year, not past month	N/A	2	N/A
Completely missing ³	None	None	None	1	1. RU1 2. RU2 3. RU3
			Past month	0	N/A
		Not past month	None	2	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	1	N/A
		Lifetime, not past year	N/A	9	N/A
		Never used	N/A	46	N/A
Total cases requiring imputation	-	-	-	61	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used PCP during lifetime and PCP recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only PCP recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of PCP | lifetime use of PCP)

2. RU2 = P(Past year but not past month use of PCP | lifetime use of PCP)

3. RU3 = P(Lifetime but not past year use of PCP | lifetime use of PCP)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.85 Imputation Details for Ecstasy Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	4	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	0	N/A
		Past year, not past month	N/A	3	N/A
Completely missing ³	None	None	None	4	1. RU1 2. RU2 3. RU3
			Past month	2	N/A
		Not past month	None	6	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	3	N/A
		Lifetime, not past year	N/A	16	N/A
		Never used	N/A	63	N/A
Total cases requiring imputation	-	-	-	101	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used Ecstasy during lifetime and Ecstasy recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only Ecstasy recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of Ecstasy | lifetime use of Ecstasy)

2. RU2 = P(Past year but not past month use of Ecstasy | lifetime use of Ecstasy)

3. RU3 = P(Lifetime but not past year use of Ecstasy | lifetime use of Ecstasy)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.86 Imputation Details for Ketamine Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	1	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	0	N/A
		Past year, not past month	N/A	1	N/A
Completely missing ³	None	None	None	2	1. RU1 2. RU2 3. RU3
			Past month	0	N/A
		Not past month	None	1	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	0	N/A
		Lifetime, not past year	N/A	15	N/A
		Never used	N/A	81	N/A
Total cases requiring imputation	-	-	-	101	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used ketamine during lifetime and ketamine recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only ketamine recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of ketamine | lifetime use of ketamine)

2. RU2 = P(Past year but not past month use of ketamine | lifetime use of ketamine)

3. RU3 = P(Lifetime but not past year use of ketamine | lifetime use of ketamine)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.87 Imputation Details for AMT/DMT/Foxy Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	2	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	0	N/A
		Past year, not past month	N/A	0	N/A
Completely missing ³	None	None	None	2	1. RU1 2. RU2 3. RU3
			Past month	0	N/A
		Not past month	None	2	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	1	N/A
		Lifetime, not past year	N/A	6	N/A
		Never used	N/A	109	N/A
Total cases requiring imputation	-	-	-	122	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used AMT/DMT/Foxy during lifetime and AMT/DMT/Foxy recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only AMT/DMT/Foxy recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of AMT/DMT/Foxy | lifetime use of AMT/DMT/Foxy)

2. RU2 = P(Past year but not past month use of AMT/DMT/Foxy | lifetime use of AMT/DMT/Foxy)

3. RU3 = P(Lifetime but not past year use of AMT/DMT/Foxy | lifetime use of AMT/DMT/Foxy)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.88 Imputation Details for *Salvia divinorum* Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Restrictions Based on Imputation-Revised Hallucinogen Recency of Use	Restrictions Based on Other Hallucinogen Child Use ¹	Number of Cases	Predictive Mean Vector ²
Past year	None	None	None	3	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
			Past month	0	N/A
		Past year, not past month	N/A	0	N/A
Completely missing ³	None	None	None	3	1. RU1 2. RU2 3. RU3
			Past month	1	N/A
		Not past month	None	1	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
			Past year, not past month	0	N/A
		Lifetime, not past year	N/A	8	N/A
		Never used	N/A	123	N/A
Total cases requiring imputation	-	-	-	139	-

N/A = not applicable.

NOTE: Other hallucinogens for recency directly include peyote, mescaline, and psilocybin and indirectly include any other hallucinogens not covered by LSD, PCP, Ecstasy, ketamine, DMT/AMT/Foxy, or *Salvia divinorum*.

¹ Respondent imputed to IRHALLUCREC, if respondent did not use any "other" hallucinogen child drugs and

1. Respondent only used *Salvia divinorum* during lifetime and *Salvia divinorum* recency is missing or

2. Respondent used multiple hallucinogen child drugs during lifetime, only *Salvia divinorum* recency is missing, and the minimum of the nonmissing recencies is greater than IRHALLUCREC.

² The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of *Salvia divinorum* | lifetime use of *Salvia divinorum*)

2. RU2 = P(Past year but not past month use of *Salvia divinorum* | lifetime use of *Salvia divinorum*)

3. RU3 = P(Lifetime but not past year use of *Salvia divinorum* | lifetime use of *Salvia divinorum*)

³ Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.89 Imputation Details for Sedatives Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation- Revised Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. $RM1/(RM1+RM2)$ 2. $RM2/(RM1+RM2)$
Lifetime	None	7	1. $RM1/(RM1+RM2+RM3)$ 2. $RM2/(RM1+RM2+RM3)$ 3. $RM3/(RM1+RM2+RM3)$
	Lifetime not past year misuse	1	N/A
Not past year	None	36	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	189	N/A
Completely missing	None	31	1. RM1 2. RM2 3. RM3 4. RM4
	Did not misuse in past year (lifetime misuse unknown)	6	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	290	N/A
Total cases requiring imputation	-	560	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. $RM1 = P(\text{Past month misuse of sedatives} \mid \text{lifetime use of sedatives})$
2. $RM2 = P(\text{Past year but not past month misuse of sedatives} \mid \text{lifetime use of sedatives})$
3. $RM3 = P(\text{Lifetime but not past year misuse of sedatives} \mid \text{lifetime use of sedatives})$
4. $RM4 = P(\text{Never misused sedatives} \mid \text{lifetime use of sedatives})$

Table E.90 Imputation Details for Stimulants Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation- Revised Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	4	1. $RM1/(RM1+RM2)$ 2. $RM2/(RM1+RM2)$
Lifetime	None	2	1. $RM1/(RM1+RM2+RM3)$ 2. $RM2/(RM1+RM2+RM3)$ 3. $RM3/(RM1+RM2+RM3)$
	Lifetime not past year misuse	0	N/A
Not past year	None	23	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	134	N/A
Completely missing	None	53	1. RM1 2. RM2 3. RM3 4. RM4
	Did not misuse in past year (lifetime misuse unknown)	6	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	237	N/A
Total cases requiring imputation	-	459	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. $RM1 = P(\text{Past month misuse of stimulants} \mid \text{lifetime use of stimulants})$
2. $RM2 = P(\text{Past year but not past month misuse of stimulants} \mid \text{lifetime use of stimulants})$
3. $RM3 = P(\text{Lifetime but not past year misuse of stimulants} \mid \text{lifetime use of stimulants})$
4. $RM4 = P(\text{Never misused stimulants} \mid \text{lifetime use of stimulants})$

Table E.91 Imputation Details for Tranquilizers Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation- Revised Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	5	1. $RM1/(RM1+RM2)$ 2. $RM2/(RM1+RM2)$
Lifetime	None	6	1. $RM1/(RM1+RM2+RM3)$ 2. $RM2/(RM1+RM2+RM3)$ 3. $RM3/(RM1+RM2+RM3)$
	Lifetime not past year misuse	1	N/A
Not past year	None	36	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	278	N/A
Completely missing	None	55	1. RM1
			2. RM2
			3. RM3
			4. RM4
	Did not misuse in past year (lifetime misuse unknown)	6	1. $RM3/(RM3+RM4)$ 2. $RM4/(RM3+RM4)$
	Never misused	236	N/A
Total cases requiring imputation	-	623	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. $RM1 = P(\text{Past month misuse of tranquilizers} \mid \text{lifetime use of tranquilizers})$
2. $RM2 = P(\text{Past year but not past month misuse of tranquilizers} \mid \text{lifetime use of tranquilizers})$
3. $RM3 = P(\text{Lifetime but not past year misuse of tranquilizers} \mid \text{lifetime use of tranquilizers})$
4. $RM4 = P(\text{Never misused tranquilizers} \mid \text{lifetime use of tranquilizers})$

Table E.92 Imputation Details for Pain Relievers Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation-Revised Recency of Any Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	14	1. RM1/(RM1+RM2) 2. RM2/(RM1+RM2)
Lifetime	None	35	1. RM1/(RM1+RM2+RM3) 2. RM2/(RM1+RM2+RM3) 3. RM3/(RM1+RM2+RM3)
	Lifetime not past year misuse	9	N/A
Not past year	None	278	1. RM3/(RM3+RM4) 2. RM4/(RM3+RM4)
	Never misused	318	N/A
Completely missing	None	117	1. RM1 2. RM2 3. RM3 4. RM4
	Did not misuse in past year (lifetime misuse unknown)	21	1. RM3/(RM3+RM4) 2. RM4/(RM3+RM4)
	Never misused	176	N/A
	-	968	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past month misuse of pain relievers | lifetime use of pain relievers)
2. RM2 = P(Past year but not past month misuse of pain relievers | lifetime use of pain relievers)
3. RM3 = P(Lifetime but not past year misuse of pain relievers | lifetime use of pain relievers)
4. RM4 = P(Never misused pain relievers | lifetime use of pain relievers)

Table E.93 Imputation Details for OxyContin Past Year Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation-Revised Recency of Any Use	Restrictions Based on Imputation-Revised Pain Relievers Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	None	5	1. RM1 2. RM2
		No past year misuse	12	N/A
	No past year misuse	N/A	154	N/A
Total cases requiring imputation	-	-	171	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past year misuse of OxyContin | past year misuse of pain relievers and past year use of OxyContin)
2. RM2 = P(No past year misuse of OxyContin | past year misuse of pain relievers and past year use of OxyContin)

Table E.94 Imputation Details for Prescription Fentanyl Past Year Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation-Revised Recency of Any Use	Restrictions Based on Imputation-Revised Pain Relievers Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	None	3	1. RM1 2. RM2
		No past year misuse	5	N/A
	No past year misuse	N/A	160	N/A
Total cases requiring imputation	-	-	168	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past year misuse of prescription fentanyl | past year misuse of pain relievers and past year use of prescription fentanyl)
2. RM2 = P(No past year misuse of prescription fentanyl | past year misuse of pain relievers and past year use of prescription fentanyl)

Table E.95 Imputation Details for Benzodiazepines Past Year Recency of Misuse

Edited Recency of Misuse	Restrictions Based on Imputation-Revised Recency of Any Use	Restrictions Based on Imputation-Revised Sedatives Recency of Misuse and Tranquilizers Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	None	8	1. RM1 2. RM2
		No past year misuse	44	N/A
	No past year misuse	N/A	283	N/A
Total cases requiring imputation	-	-	335	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past year misuse of Benzodiazepines | past year misuse use of sedatives or past year misuse use of tranquilizers, and past year use of Benzodiazepines)
2. RM2 = P(No past year misuse of Benzodiazepines | past year misuse use of sedatives or past year misuse use of tranquilizers, and past year use of Benzodiazepines)

Table E.96 Imputation Details for Nicotine or Tobacco Vaping Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. RU1/(RU1+RU2) 2. (RU2)/(RU1+RU2)
Not past month	None	116	1. RU2/(RU2+RU3+RU4) 2. RU3/(RU2+RU3+RU4) 3. RU4/(RU2+RU3+RU4)
Not past year	None	212	1. RU3/(RU3+RU4) 2. RU4/(RU3+RU4)
Past 3 years	None	7	1. RU1/(RU1+RU2+RU3) 2. RU2/(RU1+RU2+RU3) 3. RU3/(RU1+RU2+RU3)
Past 3 years but not past month	None	472	1. RU2/(RU2+RU3) 2. RU3/(RU2+RU3)
Completely missing ²	None	13	1. RU1 2. RU2 3. RU3 4. RU4
	Never used	13	N/A
Total cases requiring imputation	-	833	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of nicotine or tobacco vaping | lifetime use of nicotine or tobacco vaping)
2. RU2 = P(Past year but not past month use of nicotine or tobacco vaping | lifetime use of nicotine or tobacco vaping)
3. RU3 = P(Past 3 years but not past year use of nicotine or tobacco vaping | lifetime use of nicotine or tobacco vaping)
4. RU4 = P(Lifetime but not past 3 year use of nicotine or tobacco vaping | lifetime use of nicotine or tobacco vaping)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.97 Imputation Details for Flavoring Vaping Recency of Use

Edited Recency of Use	Restrictions Based on Imputation-Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing ²	None	236	1. RU1 2. RU2 3. RU3
	Never used	1,498	N/A
Total cases requiring imputation	-	1,734	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of flavoring vaping | lifetime use of flavoring vaping)
2. RU2 = P(Past year but not past month use of flavoring vaping | lifetime use of flavoring vaping)
3. RU3 = P(Lifetime but not past year use of flavoring vaping | lifetime use of flavoring vaping)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.98 Imputation Details for CBD or Hemp Recency of Use

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Completely missing ²	None	121	1. RU1 2. RU2 3. RU3 4. RU4
	Never used	451	N/A
Total cases requiring imputation	-	572	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of CBD or hemp | lifetime use of CBD or hemp)
2. RU2 = P(Past year but not past month use of CBD or hemp | lifetime use of CBD or hemp)
3. RU3 = P(Past 3 years but not past year use of CBD or hemp | lifetime use of CBD or hemp)
4. RU4 = P(Lifetime but not past 3 year use of CBD or hemp | lifetime use of CBD or hemp)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.99 Imputation Details for Illegally Made Fentanyl Recency of Use

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	21	1. RU1 2. RU2 3. RU3
	Never used	1,602	N/A
Total cases requiring imputation	-	1,623	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of illegally made fentanyl | lifetime use of illegally made fentanyl)
2. RU2 = P(Past year but not past month use of illegally made fentanyl | lifetime use of illegally made fentanyl)
3. RU3 = P(Lifetime but not past year use of illegally made fentanyl | lifetime use of illegally made fentanyl)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.100 Imputation Details for Kratom Recency of Use

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	46	1. RU1 2. RU2 3. RU3
	Never used	1,769	N/A
Total cases requiring imputation	-	1,815	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of kratom | lifetime use of kratom)
2. RU2 = P(Past year but not past month use of kratom | lifetime use of kratom)
3. RU3 = P(Lifetime but not past year use of kratom | lifetime use of kratom)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.**Table E.101 Imputation Details for Synthetic Marijuana Recency of Use**

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	56	1. RU1 2. RU2 3. RU3
	Never used	1,602	N/A
Total cases requiring imputation	-	1,658	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of synthetic marijuana | lifetime use of synthetic marijuana)
2. RU2 = P(Past year but not past month use of synthetic marijuana | lifetime use of synthetic marijuana)
3. RU3 = P(Lifetime but not past year use of synthetic marijuana | lifetime use of synthetic marijuana)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.

Table E.102 Imputation Details for Synthetic Stimulants Recency of Use

Edited Recency of Use	Restrictions Based on Imputation- Revised Lifetime Use	Number of Cases	Predictive Mean Vector ¹
Past year	None	0	1. RU1/(RU1+RU2) 2. RU2/(RU1+RU2)
Completely missing ²	None	12	1. RU1 2. RU2 3. RU3
	Never used	1,594	N/A
Total cases requiring imputation	-	1,606	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RU1 = P(Past month use of synthetic stimulants | lifetime use of synthetic stimulants)
2. RU2 = P(Past year but not past month use of synthetic stimulants | lifetime use of synthetic stimulants)
3. RU3 = P(Lifetime but not past year use of synthetic stimulants | lifetime use of synthetic stimulants)

² Respondents who were logically assigned lifetime edited recency of use values are included in the completely missing row.**Table E.103 Imputation Details for GHB Recency of Use**

Edited Recency of Use	Number of Cases	Predictive Mean Vector ¹
Past year	0	1. RM1/(RM1+RM2) 2. RM2/(RM1+RM2)
Lifetime	4	1. RM1/(RM1+RM2+RM3) 2. RM2/(RM1+RM2+RM3) 3. RM3/(RM1+RM2+RM3)
Completely missing	347	1. RM1 2. RM2 3. RM3 4. RM4
Total cases requiring imputation	351	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past month use of GHB)
2. RM2 = P(Past year but not past month use of GHB)
3. RM3 = P(Lifetime but not past year use of GHB)
4. RM4 = P(Never used GHB)

Table E.104 Imputation Details for Nonprescription Cough or Cold Medicine Recency of Use

Edited Recency of Use	Number of Cases	Predictive Mean Vector ¹
Past year	0	1. RM1/(RM1+RM2) 2. RM2/(RM1+RM2)
Lifetime	55	1. RM1/(RM1+RM2+RM3) 2. RM2/(RM1+RM2+RM3) 3. RM3/(RM1+RM2+RM3)
Completely missing	395	1. RM1 2. RM2 3. RM3 4. RM4
Total cases requiring imputation	450	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. RM1 = P(Past month use of nonprescription cough or cold medicine)
2. RM2 = P(Past year but not past month use of nonprescription cough or cold medicine)
3. RM3 = P(Lifetime but not past year use of nonprescription cough or cold medicine)
4. RM4 = P(Never used nonprescription cough or cold medicine)

E.3.2.2.3 12-Month Frequency of Use**Table E.105 Logical Constraints for Inhalants and Methamphetamine 12-Month Frequency**

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be within the range defined by the recipient's minimum and maximum possible substance frequency of use in the past year Recipient's maximum number of days could have used in the past year is determined by the month of first substance use Calculation of recipient's minimum and maximum possible substance frequency of use in the past year is described in Chapter 6

Table E.106 Likeness Constraints for Inhalants and Methamphetamine 12-Month Frequency

Constraint #	Likeness Constraint
LikC1	Donor's substance recency must match recipient's substance recency
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor and recipient are in the same response propensity bin

Table E.107 Constraints and Portion of the Predictive Mean Vector for Inhalants 12-Month Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	74	1	1. PY	No Cases	1-4: 3 1,2,4: 2	1-4: 3 1,2,4: 6	1-4: 6 1,2,4: 7	1-4: 14 1,2,4: 33
Cycle 2									
1	Missing	74	1	1. PY	No Cases	1-4: 1 1,2,4: 4	1-4: 3 1,2,4: 6	1-4: 5 1,2,4: 8	1-4: 13 1,2,4: 34

¹ The predictive mean vector components are defined by the following:
1. PY = P(Inhalant use on a given day in the past year | past year inhalant use)

Table E.108 Constraints and Portion of the Predictive Mean Vector for Methamphetamine 12-Month Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	15	1	1. PY	No Cases	1-4: 0 1,2,4: 1	1-4: 0 1,2,4: 2	1-4: 0 1,2,4: 3	1-4: 2 1,2,4: 7
Cycle 2									
1	Missing	15	1	1. PY	No Cases	1-4: 1 1,2,4: 2	1-4: 1 1,2,4: 1	1-4: 0 1,2,4: 4	1-4: 0 1,2,4: 6

¹ The predictive mean vector components are defined by the following:
1. PY = P(Methamphetamine use on a given day in the past year | past year methamphetamine use)

Table E.109 Logical Constraints for Hallucinogens 12-Month Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be less than or equal to recipient's maximum possible past year substance frequency of use Recipient's maximum possible substance frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used substance, as determined by the month of first substance use (2) if the maximum period recipient could have used substance is greater than 30, but recipient is a past month substance user with a nonmissing 30-day substance frequency, then the past year substance frequency must be less than or equal to the maximum period (minus the number of days recipient did not use substance in the past month) (3) if recipient is not a past month substance user, then the past year substance frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year substance use * recipient's maximum number of days could have used substance in the past year must be greater than or equal to recipient's minimum possible past year substance frequency of use Recipient's minimum possible substance frequency of use in the past year is limited by the following factors: (1) if recipient is a past month substance user, it must be at least as much as the 30-day substance frequency (2) if recipient is not a past month substance user but is a past year substance user, it must be at least 1
LogC3	If recipient's age at first use equals recipient's current age, and (1) donor is a past month hallucinogen user, then donor's proportion of past year use * recipient's maximum number of days could have used in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive) (2) donor is a past year but not past month user, then donor's proportion of past year use * recipient's maximum number of days could have used in the past year must be less than or equal to the number of days between recipient's interview date and birthday (minus 30)

Table E.110 Likeness Constraints for Hallucinogens 12-Month Frequency

Constraint #	Likeness Constraint
LikC1	Donor's substance recency must match recipient's substance recency
LikC2	State rank of donor = state rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor and recipient are in the same response propensity bin

Table E.111 Constraints and Portion of the Predictive Mean Vector for Hallucinogens 12-Month Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	118	1,2,3	1. PY	1-4: 6 1,2,4: 3	1-4: 14	1-4: 18 1,2,4: 4	1-4: 17 1,2,4: 6	1-4: 34 1,2,4: 16
Cycle 2									
1	Missing	118	1,2,3	1. PY	1-4: 7 1,2,4: 2	1-4: 12 1,2,4: 3	1-4: 17 1,2,4: 2	1-4: 19 1,2,4: 4	1-4: 41 1,2,4: 11

¹ The predictive mean vector components are defined by the following:

1. PY = P(Hallucinogen use on a given day in the past year | past year hallucinogen use)

*E.3.2.2.4 30-Day Frequency of Use***Table E.112 Logical Constraints for Nicotine Vaping, Inhalants, Methamphetamine, Hallucinogens, Sedatives, Stimulants, Tranquilizers, and Pain Relievers 30-Day Frequency**

Constraint #	Logical Constraint
LogC1	Donor's 30-day substance use must be less than or equal to recipient's maximum number of days could have used substance in past 30 days and greater than or equal to recipient's minimum number of days could have used substance in past 30 days

Table E.113 Likeness Constraints for Inhalants, Methamphetamine, Hallucinogens, Sedatives, Stimulants, Tranquilizers, and Pain Relievers 30-Day Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.114 Constraints and Portion of the Predictive Mean Vector for Nicotine Vaping 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	51	1	1. PM	No Cases	1,2: 1	1,2: 5	1,2: 8	1,2: 37

¹ The predictive mean vector components are defined by the following:
 1. PM = P(Nicotine vaping on a given day in the past month | past month nicotine vaping)

Table E.115 Constraints and Portion of the Predictive Mean Vector for Inhalants 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	21	1	1. PM	No Cases	1,2: 1	1,2: 3 1: 2	1,2: 2 1: 3	1,2: 6 1: 4
Cycle 2									
1	Missing	21	1	1. PM	1,2: 1	1,2: 1 1: 1	1,2: 2	1,2: 3	1,2: 8 1: 5

¹ The predictive mean vector components are defined by the following:
 1. PM = P(Inhalant use on a given day in the past month | past month inhalant use)

Table E.116 Constraints and Portion of the Predictive Mean Vector for Methamphetamine 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	12	1	1. PM	1,2: 1	1,2: 0 1: 1	1,2: 0 1: 1	1,2: 0 1: 2	1,2: 2 1: 5
Cycle 2									
1	Missing	12	1	1. PM	No Cases	1,2: 1 1: 1	1,2: 2	1,2: 0 1: 1	1,2: 4 1: 3

¹ The predictive mean vector components are defined by the following:

1. PM = P(Methamphetamine use on a given day in the past month | past month methamphetamine use)

Table E.117 Constraints and Portion of the Predictive Mean Vector for Hallucinogens 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	69	1	1. PM	1,2: 5	1,2: 9 1: 2	1,2: 10 1: 2	1,2: 11 1: 1	1,2: 22 1: 7
Cycle 2									
1	Missing	69	1	1. PM	1,2: 4 1: 1	1,2: 11	1,2: 13	1,2: 16 1: 2	1,2: 17 1: 5

¹ The predictive mean vector components are defined by the following:

1. PM = P(Hallucinogens use on a given day in the past month | past month hallucinogens use)

Table E.118 Constraints and Portion of the Predictive Mean Vector for Sedatives 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	1	1	1. PM	1,2: 0 1: 1	No Cases	No Cases	No Cases	No Cases
Cycle 2									
1	Missing	1	1	1. PM	1,2: 0 1: 1	No Cases	No Cases	No Cases	No Cases

¹ The predictive mean vector components are defined by the following:
 1. PM = P(Sedative use on a given day in the past month | past month sedative use)

Table E.119 Constraints and Portion of the Predictive Mean Vector for Stimulants 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	2	1	1. PM	1,2: 0 1: 1	No Cases	1,2: 1	No Cases	No Cases
Cycle 2									
1	Missing	2	1	1. PM	1,2: 0 1: 1	1,2: 0 1: 1	No Cases	No Cases	No Cases

¹ The predictive mean vector components are defined by the following:
 1. PM = P(Stimulant use on a given day in the past month | past month stimulant use)

Table E.120 Constraints and Portion of the Predictive Mean Vector for Tranquilizers 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	4	1	1. PM	No Cases	No Cases	1,2: 1	No Cases	1,2: 1 1: 2
Cycle 2									
1	Missing	4	1	1. PM	No Cases	No Cases	No Cases	1,2: 0 1: 1	1,2: 0 1: 3

¹ The predictive mean vector components are defined by the following:

1. PM = P(Tranquilizer use on a given day in the past month | past month tranquilizer use)

Table E.121 Constraints and Portion of the Predictive Mean Vector for Pain Relievers 30-Day Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	10	1	1. PM	No Cases	1,2: 1	No Cases	1,2: 0 1: 1	1,2: 5 1: 3
Cycle 2									
1	Missing	10	1	1. PM	1,2: 0 1: 1	No Cases	1,2: 0 1: 1	1,2: 1	1,2: 7

¹ The predictive mean vector components are defined by the following:

1. PM = P(Pain reliever use on a given day in the past month | past month pain reliever use)

*E.3.2.2.5 Past Year Initiation of Misuse***Table E.122 Imputation Details for Past Year Initiation of Misuse for Sedatives**

Edited Past Year Initiation of Misuse Indicator	Restrictions Based on Imputation-Revised Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	6	1. INIT1 2. INIT2
	Not a past year initiate	329	N/A
Total cases requiring imputation	-	335	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. INIT1 = P(Past year initiation of sedatives | past year misuse of sedatives)
2. INIT2 = P(No past year initiation of sedatives | past year misuse of sedatives)

Table E.123 Imputation Details for Past Year Initiation of Misuse for Stimulants

Edited Past Year Initiation of Misuse Indicator	Restrictions Based on Imputation-Revised Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	11	1. INIT1 2. INIT2
	Not a past year initiate	293	N/A
Total cases requiring imputation	-	304	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. INIT1 = P(Past year initiation of stimulants | past year misuse of stimulants)
2. INIT2 = P(No past year initiation of stimulants | past year misuse of stimulants)

Table E.124 Imputation Details for Past Year Initiation of Misuse for Tranquilizers

Edited Past Year Initiation of Misuse Indicator	Restrictions Based on Imputation-Revised Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	14	1. INIT1 2. INIT2
	Not a past year initiate	293	N/A
Total cases requiring imputation	-	307	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. INIT1 = P(Past year initiation of tranquilizers | past year misuse of tranquilizers)
2. INIT2 = P(No past year initiation of tranquilizers | past year misuse of tranquilizers)

Table E.125 Imputation Details for Past Year Initiation of Misuse for Pain Relievers

Edited Past Year Initiation of Misuse Indicator	Restrictions Based on Imputation-Revised Recency of Misuse	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	42	1. INIT1 2. INIT2
	Not a past year initiate	332	N/A
Total cases requiring imputation	-	374	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. INIT1 = P(Past year initiation of pain relievers | past year misuse of pain relievers)
2. INIT2 = P(No past year initiation of pain relievers | past year misuse of pain relievers)

E.3.2.3 Age at First Use

Table E.126 Logical Constraints for modPMN Age at First Use, Univariate

Constraint #	Logical Constraint
LogC1	Donor's age at first substance use must be less than or equal to recipient's age
LogC2	If recipient was not a past year substance user, then donor cannot have age at first substance use equal to recipient's age
LogC3	If recipient's 12-month substance frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first substance use equal to recipient's age
LogC4	If recipient's most recent birthday was within the past 30 days, and recipient's substance recency was past year but not past month, then donor cannot have age at first substance use equal to recipient's age
LogC5	If recipient's substance recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month substance frequency was less than 30, then donor cannot have age at first substance use equal to recipient's age. (These recipients are missed by LogC3, but the idea is the same: the date of first substance use must be earlier than recipient's most recent birthday.)
LogC6	If recipient was not a past year user of hallucinogens or was not a past year user of any hallucinogen child drugs (LSD, PCP, Ecstasy, ketamine, <i>Salvia divinorum</i> , or DMT/AMT/Foxy), then donor cannot have hallucinogen age at first drug use equal to recipient's age
LogC7	If recipient's most recent birthday was within the past 30 days, and recipient's hallucinogen recency was past year but not past month or any hallucinogen child drugs recency (LSD, PCP, Ecstasy, ketamine, <i>Salvia divinorum</i> , or DMT/AMT/Foxy) was past year but not past month, then donor cannot have hallucinogen age at first drug use equal to recipient's age
LogC8	If recipient was missing the age at first hallucinogen use, then donor must have hallucinogen age at first use less than or equal to the recipient's child drug(s) age(s) at first use (if existing; LSD, PCP, or Ecstasy)
LogC9	If recipient was not a lifetime user of "other" ¹ hallucinogens, then donor must not have values that would cause the minimum of the child drug(s) age(s) at first use to be greater than the overall hallucinogens age at first use for recipient. For example, if donor was not a lifetime user of "other" hallucinogens and has HALLUCAGE = 17, LSDAGE = missing, PCPAGE = 19, and ECSTMOAGE = missing, then donor must have LSDAGE = 17 and/or ECSTMOAGE = 17.
LogC10	Donor's age at first use must be greater than or equal to recipient's hallucinogen age at first use
LogC11	If recipient did not vape nicotine in the past 3 years, then donor's age at first use must be at least 3 years less than recipient's age

¹ Other hallucinogens for LogC12 directly include peyote, mescaline, psilocybin, ketamine, DMT/AMT/Foxy, *Salvia divinorum* and indirectly include any other hallucinogens indicated by write-in responses.

Table E.127 Likeness Constraints for modPMN Age at First Use, Univariate

Constraint #	Likeness Constraint
LikC1	State rank of donor = state rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = age of recipient
LikC4	Age of donor must be greater than or equal to age of recipient
LikC5	If recipient was a past year substance user, then donor was a past year substance user; if recipient was not a past year substance user, then donor was the same
LikC6	If recipient was not a past year substance user, then donor was the same
LikC7	If recipient was a past year user, then donor was a past year user
LikC8	For each child drug (LSD, PCP, Ecstasy), if recipient was a past year user, then donor was a past year user
LikC9	For each child drug (LSD, PCP, Ecstasy), if recipient was a lifetime but not past year user, then donor was not a past year user
LikC10	Donor agrees with recipient with respect to lifetime use for child drug(s) (LSD, PCP, Ecstasy)
LikC11	Donor was at least as old as recipient, but no more than 20 years older than recipient
LikC12	Donor was no more than 20 years older than recipient
LikC13	Donor and recipient are in the same response propensity bin

Table E.128 Constraints and Portion of the Predictive Mean Vector for Nicotine Vaping Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	460	1,2,7,11	1. PrAFU	1-3,5,13: 17	1-3,5,13: 14	1-3,5,13: 42	1-3,5,13: 90	1-3,5,13: 273
						2,3,5,13: 2	2,3,5,13: 2	2,3,5,13: 5	2,3,5,13: 5
						3,5,13: 1	3,5,13: 1	3,5,13: 0	3,5,13: 2
						3,6,13: 0	3,6,13: 1	3,6,13: 0	3,6,13: 3
						4,6,13: 1		4,6,13: 1	

¹ The predictive mean vector components are defined by the following:

1. PrAFU = Predicted nicotine vaping age at first use | vaped nicotine in lifetime

Table E.129 Constraints and Portion of the Predictive Mean Vector for Inhalants Age at First Use

#		Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
						Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1										
1	Missing	401	1,2,3,4,5	1. PrAFU	1-3,5,13: 13	1-3,5,13: 16	1-3,5,13: 29	1-3,5,13: 65	1-3,5,13: 190	
					2,3,5,13: 2	2,3,5,13: 2	2,3,5,13: 10	2,3,5,13: 9	2,3,5,13: 15	
					3,5,13: 3	3,5,13: 0	3,5,13: 6	3,5,13: 8	3,5,13: 27	
						3,6,13: 0	3,6,13: 1	3,6,13: 1	3,6,13: 1	
						4,6,13: 1			4,6,13: 2	
Cycle 2										
1	Missing	401	1,2,3,4,5	1. PrAFU	1-3,5,13: 15	1-3,5,13: 16	1-3,5,13: 25	1-3,5,13: 70	1-3,5,13: 202	
					2,3,5,13: 2	2,3,5,13: 2	2,3,5,13: 11	2,3,5,13: 5	2,3,5,13: 9	
					3,5,13: 1	3,5,13: 0	3,5,13: 9	3,5,13: 7	3,5,13: 21	
						3,6,13: 0	3,6,13: 1	3,6,13: 1	3,6,13: 1	
						4,6,13: 1			4,6,13: 2	

¹ The predictive mean vector components are defined by the following:

1. PrAFU = Predicted inhalant age at first use | used inhalants in lifetime

Table E.130 Constraints and Portion of the Predictive Mean Vector for Methamphetamine Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	21	1,2,3,4,5	1. PrAFU	No Cases	1-3,5,13: 1	1-3,5,13: 0	1-3,5,13: 1	1-3,5,13: 4
							2,3,5,13: 0	2,3,5,13: 1	2,3,5,13: 1
							3,5,13: 2	3,5,13: 1	3,5,13: 9
							3,6,13: 0		
							4,6,13: 1		
Cycle 2									
1	Missing	21	1,2,3,4,5	1. PrAFU	No Cases	1-3,5,13: 1	1-3,5,13: 0	1-3,5,13: 3	1-3,5,13: 5
							2,3,5,13: 1		2,3,5,13: 2
							3,5,13: 1		3,5,13: 7
							3,6,13: 0		
							4,6,13: 1		

¹ The predictive mean vector components are defined by the following:

1. PrAFU = Predicted methamphetamine age at first use | used methamphetamine in lifetime

Table E.131 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	159	1,3,5-9	1. PrHALAFU	1-3,6-10,13: 3	1-3,6-10,13: 4	1-3,6-10,13: 5	1-3,6-10,13: 8	1-3,6-10,13: 28
					2,3,6-10,13: 1	2,3,6-10,13: 4	2,3,6-10,13: 3	2,3,6-10,13: 5	
					2,3,6-9,13: 5	2,3,6-9,13: 3	2,3,6-9,13: 5	2,3,6-9,13: 8	2,3,6-10,13: 22
							3,6-9,13: 2	3,6-9,13: 6	
							3,6,9,13: 0	3,6,9,13: 1	2,3,6-9,13: 17
							6,9,11,13: 1	6,9,11,13: 1	3,6-9,13: 20
								6,9,12,13: 1	3,6,9,13: 3
									6,9,11,13: 2
									6,9,12,13: 1

Table E.131 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Age at First Use (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 2									
1	Missing	159	1,3,5-9	1. PrHALAFU	1-3,6-10,13: 3	1-3,6-10,13: 4	1-3,6-10,13: 5	1-3,6-10,13: 10	1-3,6-10,13: 26
					2,3,6-10,13: 2	2,3,6-10,13: 6	2,3,6-10,13: 3	2,3,6-10,13: 4	2,3,6-10,13: 22
					2,3,6-9,13: 3	2,3,6-9,13: 1	2,3,6-9,13: 3	2,3,6-9,13: 9	2,3,6-9,13: 17
					3,6-9,13: 1		3,6-9,13: 4	3,6-9,13: 4	3,6-9,13: 22
							3,6,9,13: 0	3,6,9,13: 1	3,6,9,13: 3
							6,9,11,13: 1	6,9,11,13: 1	6,9,11,13: 2
									6,9,12,13: 1

¹ The predictive mean vector components are defined by the following:

1. PrHALAFU = Predicted hallucinogen age at first use | used hallucinogens in lifetime

Table E.132 Constraints and Portion of the Predictive Mean Vector for LSD Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	70	1,2,4,10	1. PrLSDAFU	1-3,6,7,13: 4	1-3,6,7,13: 1	1-3,6,7,13: 5	1-3,6,7,13: 6	1-3,6,7,13: 6
					2,3,6,7,13: 0	2,3,6,7,13: 2	2,3,6,7,13: 3	2,3,6,7,13: 3	2,3,6,7,13: 10
					3,6,7,13: 1	3,6,7,13: 1	3,6,7,13: 1	3,6,7,13: 5	3,6,7,13: 15
						3,6,13: 0			3,6,13: 2
					6,11,13: 1			6,11,13: 4	

Table E.132 Constraints and Portion of the Predictive Mean Vector for LSD Age at First Use (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 2									
					1-3,6,7,13: 1	1-3,6,7,13: 2	1-3,6,7,13: 4	1-3,6,7,13: 6	1-3,6,7,13: 11
					2,3,6,7,13: 1	2,3,6,7,13: 2	2,3,6,7,13: 1	2,3,6,7,13: 4	2,3,6,7,13: 5
						3,6,7,13: 1	3,6,7,13: 1	3,6,7,13: 8	3,6,7,13: 14
							3,6,13: 0		3,6,13: 2
							6,11,13: 2		6,11,13: 4
1	Missing	70	1,2,4,10	1. PrLSDAFU			6,12,13: 0		
							12,13: 0		
							13: 0		
							1-3,6,7: 0		
							2,3,6,7: 0		
							3,6,7: 1		

¹ The predictive mean vector components are defined by the following:

1. PrLSDAFU = Predicted LSD age at first use | used LSD in lifetime

Table E.133 Constraints and Portion of the Predictive Mean Vector for PCP Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	28	1,2,4,10	1. PrPCPAFU	1-3,6,7,13: 0	1-3,6,7,13: 0	1-3,6,7,13: 0	1-3,6,7,13: 0	1-3,6,7,13: 0
					2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 1
					3,6,7,13: 1	3,6,7,13: 3	3,6,7,13: 3	3,6,7,13: 4	3,6,7,13: 5
					3,6,13: 0		3,6,13: 0	3,6,13: 0	3,6,13: 0
					6,11,13: 2		6,11,13: 1	6,11,13: 1	6,11,13: 7
Cycle 2									
1	Missing	28	1,2,4,10	1. PrPCPAFU	1-3,6,7,13: 0	1-3,6,7,13: 1	1-3,6,7,13: 0	1-3,6,7,13: 1	1-3,6,7,13: 1
					2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 0	2,3,6,7,13: 0
					3,6,7,13: 0	3,6,7,13: 1	3,6,7,13: 5	3,6,7,13: 3	3,6,7,13: 6
					3,6,13: 0		3,6,13: 0	3,6,13: 1	3,6,13: 0
					6,11,13: 1		6,11,13: 1	6,11,13: 1	6,11,13: 6

¹ The predictive mean vector components are defined by the following:

1. PrPCPAFU = Predicted PCP age at first use | used PCP in lifetime

Table E.134 Constraints and Portion of the Predictive Mean Vector for Ecstasy Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Cycle 1									
1	Missing	32	1,2,4,10	1. PrECSAFU	1-3,6,7,13: 3	1-3,6,7,13: 2	1-3,6,7,13: 2	1-3,6,7,13: 2	1-3,6,7,13: 8
								2,3,6,7,13: 2	2,3,6,7,13: 3
								3,6,7,13: 3	3,6,7,13: 4
									3,6,13: 1
									6,11,13: 1
				6,12,13: 1					
Cycle 2									
1	Missing	32	1,2,4,10	1. PrECSAFU	1-3,6,7,13: 3	1-3,6,7,13: 2	1-3,6,7,13: 1	1-3,6,7,13: 3	1-3,6,7,13: 7
							2,3,6,7,13: 1	2,3,6,7,13: 2	2,3,6,7,13: 2
								3,6,7,13: 2	3,6,7,13: 6
									3,6,13: 1
									6,11,13: 1
				6,12,13: 1					

¹ The predictive mean vector components are defined by the following:
1. PrAFU = Predicted Ecstasy age at first use | used Ecstasy in lifetime

Table E.135 Imputation Details for Age of First Misuse for Sedatives

Edited Age of First Misuse	Restrictions Based on Imputation-Revised Past Year Initiation Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	10	1. AFM1 2. AFM2
	Legitimate skip - not a past year initiate	333	N/A
Total cases requiring imputation	-	343	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. AFM1 = P(First misuse of sedatives at current age | past year initiation of sedatives)
2. AFM2 = P(First misuse of sedatives at age – 1 | past year initiation of sedatives)

Table E.136 Imputation Details for Age of First Misuse for Stimulants

Edited Age of First Misuse	Restrictions Based on Imputation-Revised Past Year Initiation Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	17	1. AFM1 2. AFM2
	Legitimate skip - not a past year initiate	304	N/A
Total cases requiring imputation	-	321	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. AFM1 = P(First misuse of stimulants at current age | past year initiation of stimulants)
2. AFM2 = P(First misuse of stimulants at age – 1 | past year initiation of stimulants)

Table E.137 Imputation Details for Age of First Misuse for Tranquilizers

Edited Age of First Misuse	Restrictions Based on Imputation-Revised Past Year Initiation Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	14	1. AFM1 2. AFM2
	Legitimate skip - not a past year initiate	305	N/A
Total cases requiring imputation	-	319	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. AFM1 = P(First misuse of tranquilizers at current age | past year initiation of tranquilizers)
2. AFM2 = P(First misuse of tranquilizers at age – 1 | past year initiation of tranquilizers)

Table E.138 Imputation Details for Age of First Misuse for Pain Relievers

Edited Age of First Misuse	Restrictions Based on Imputation-Revised Past Year Initiation Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	111	1. AFM1 2. AFM2
	Legitimate skip - not a past year initiate	362	N/A
Total cases requiring imputation		473	-

N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. AFM1 = P(First misuse of pain relievers at current age | past year initiation of pain relievers)
2. AFM2 = P(First misuse of pain relievers at age – 1 | past year initiation of pain relievers)

E.3.2.4 Substance Use Disorder: DSM-5**Table E.139 Imputation Details for Past Year Alcohol Use Disorder: DSM-5**

Edited Past Year Use Disorder	Number of Cases	Predictive Mean Vector ¹
Completely missing	631	1. SUD1 2. SUD2
Total cases requiring imputation		631

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year alcohol use disorder | past year use of alcohol)
2. SUD2 = P(No past year alcohol use disorder | past year use of alcohol)

Table E.140 Imputation Details for Past Year Alcohol Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	28	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	19	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	153	1. SEV1 2. SEV2 3. SEV3
	No use disorder	486	N/A
Total cases requiring imputation		686	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year alcohol use disorder mild severity | past year alcohol use disorder)
2. SEV2 = P(Past year alcohol use disorder moderate severity | past year alcohol use disorder)
3. SEV3 = P(Past year alcohol use disorder severe severity | past year alcohol use disorder)

Table E.141 Imputation Details for Past Year Marijuana Use Disorder: DSM-5

Edited Past Year Use Disorder	Number of Cases	Predictive Mean Vector ¹
Completely missing	290	1. SUD1 2. SUD2
Total cases requiring imputation	290	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year marijuana use disorder | past year use of marijuana)
2. SUD2 = P(No past year marijuana use disorder | past year use of marijuana)

Table E.142 Imputation Details for Past Year Marijuana Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	24	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	13	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	150	1. SEV1 2. SEV2 3. SEV3
	No use disorder	150	N/A
Total cases requiring imputation	-	337	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year marijuana use disorder mild severity | past year marijuana use disorder)
2. SEV2 = P(Past year marijuana use disorder moderate severity | past year marijuana use disorder)
3. SEV3 = P(Past year marijuana use disorder severe severity | past year marijuana use disorder)

Table E.143 Imputation Details for Past Year Cocaine Use Disorder: DSM-5

Edited Past Year Use Disorder	Number of Cases	Predictive Mean Vector ¹
Completely missing	37	1. SUD1 2. SUD2
Total cases requiring imputation	37	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year cocaine use disorder | past year use of cocaine)
2. SUD2 = P(No past year cocaine use disorder | past year use of cocaine)

Table E.144 Imputation Details for Past Year Cocaine Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	1	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	0	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	8	1. SEV1 2. SEV2 3. SEV3
	No use disorder	30	N/A
Total cases requiring imputation	-	39	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year cocaine use disorder mild severity | past year cocaine use disorder)
2. SEV2 = P(Past year cocaine use disorder moderate severity | past year cocaine use disorder)
3. SEV3 = P(Past year cocaine use disorder severe severity | past year cocaine use disorder)

Table E.145 Imputation Details for Past Year Methamphetamine Use Disorder: DSM-5

Edited Past Year Use Disorder	Number of Cases	Predictive Mean Vector ¹
Completely missing	21	1. SUD1 2. SUD2
Total cases requiring imputation	21	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year methamphetamine use disorder | past year use of methamphetamine or needle use to inject methamphetamine)
2. SUD2 = P(No past year methamphetamine use disorder | past year use of methamphetamine or needle use to inject methamphetamine)

Table E.146 Imputation Details for Past Year Methamphetamine Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	1	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	0	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	13	1. SEV1 2. SEV2 3. SEV3
	No use disorder	8	N/A
Total cases requiring imputation	-	22	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year methamphetamine use disorder mild severity | past year methamphetamine use disorder)
2. SEV2 = P(Past year methamphetamine use disorder moderate severity | past year methamphetamine use disorder)
3. SEV3 = P(Past year methamphetamine use disorder severe severity | past year methamphetamine use disorder)

Table E.147 Imputation Details for Past Year Stimulant Use Disorder: DSM-5

Edited Past Year Use Disorder	Imputation-Revised Past Year Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	Past year misuse	24	1. SUD1 2. SUD2
Completely missing	Past year use but not misuse	141	1. UNM1 2. UNM2
Total cases requiring imputation	-	165	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year stimulant use disorder | past year misuse of stimulants)
2. SUD2 = P(No past year stimulant use disorder | past year misuse of stimulants)
1. UNM1 = P(Past year stimulant use disorder | past year use but not misuse of stimulants)
2. UNM2 = P(No past year stimulant use disorder | past year use but not misuse of stimulants)

Table E.148 Imputation Details for Past Year Stimulant Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Imputation-Revised Past Year Use	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	Past year misuse	None	0	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	Past year misuse	None	1	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	Past year misuse	None	9	1. SEV1 2. SEV2 3. SEV3
		No disorder	18	N/A
Mild or moderate severity	Past year use but not misuse	None	2	1. UNMS1/(UNMS1+UNMS2) 2. UNMS2/(UNMS1+UNMS2)
Moderate or severe severity	Past year use but not misuse	None	0	1. UNMS2/(UNMS2+UNMS3) 2. UNMS3/(UNMS2+UNMS3)
Completely missing severity	Past year use but not misuse	None	17	1. UNMS1 2. UNMS2 3. UNMS3
		No disorder	126	N/A
Total cases requiring imputation	-	-	173	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year stimulant use disorder mild severity | past year misuse of stimulants and stimulant use disorder)
2. SEV2 = P(Past year stimulant use disorder moderate severity | past year misuse of stimulants and stimulant use disorder)
3. SEV3 = P(Past year stimulant use disorder severe severity | past year misuse of stimulants and stimulant use disorder)
1. UNMS1 = P(Past year stimulant use disorder mild severity | past year use but not misuse of stimulants and stimulant use disorder)
2. UNMS2 = P(Past year stimulant use disorder moderate severity | past year use but not misuse of stimulants and stimulant use disorder)
3. UNMS3 = P(Past year stimulant use disorder severe severity | past year use but not misuse of stimulants and stimulant use disorder)

Table E.149 Imputation Details for Past Year Heroin Use Disorder: DSM-5

Edited Past Year Use Disorder	Number of Cases	Predictive Mean Vector ¹
Completely missing	5	1. SUD1 2. SUD2
Total cases requiring imputation	5	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year heroin use disorder | past year use of heroin)
2. SUD2 = P(No past year heroin use disorder | past year use of heroin)

Table E.150 Imputation Details for Past Year Heroin Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	0	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	0	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	5	1. SEV1 2. SEV2 3. SEV3
	No use disorder	1	N/A
Total cases requiring imputation	-	6	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year heroin use disorder mild severity | past year heroin use disorder)
2. SEV2 = P(Past year heroin use disorder moderate severity | past year heroin use disorder)
3. SEV3 = P(Past year heroin use disorder severe severity | past year heroin use disorder)

Table E.151 Imputation Details for Past Year Pain Reliever Use Disorder: DSM-5

Edited Past Year Use Disorder	Imputation-Revised Past Year Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	Past year misuse	87	1. SUD1 2. SUD2
Completely missing	Past year use but not misuse	362	1. UNM1 2. UNM2
Total cases requiring imputation	-	449	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year pain reliever use disorder | past year misuse of pain relievers)
2. SUD2 = P(No past year pain reliever use disorder | past year misuse of pain relievers)
1. UNM1 = P(Past year pain reliever use disorder | past year use but not misuse of pain relievers)
2. UNM2 = P(No past year pain reliever use disorder | past year use but not misuse of pain relievers)

Table E.152 Imputation Details for Past Year Pain Reliever Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Imputation-Revised Past Year Use	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	Past year misuse	None	2	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	Past year misuse	None	4	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	Past year misuse	None	24	1. SEV1 2. SEV2 3. SEV3
		No disorder	65	N/A
Mild or moderate severity	Past year use but not misuse	None	4	1. UNMS1/(UNMS1+UNMS2) 2. UNMS2/(UNMS1+UNMS2)
Moderate or severe severity	Past year use but not misuse	None	0	1. UNMS2/(UNMS2+UNMS3) 2. UNMS3/(UNMS2+UNMS3)
Completely missing severity	Past year use but not misuse	None	48	1. UNMS1 2. UNMS2 3. UNMS3
		No disorder	321	N/A
Total cases requiring imputation	-	-	468	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year pain reliever use disorder mild severity | past year misuse of pain relievers and pain reliever use disorder)
2. SEV2 = P(Past year pain reliever use disorder moderate severity | past year misuse of pain relievers and pain reliever use disorder)
3. SEV3 = P(Past year pain reliever use disorder severe severity | past year misuse of pain relievers and pain reliever use disorder)
1. UNMS1 = P(Past year pain reliever use disorder mild severity | past year use but not misuse of pain relievers and pain reliever use disorder)
2. UNMS2 = P(Past year pain reliever use disorder moderate severity | past year use but not misuse of pain relievers and pain reliever use disorder)
3. UNMS3 = P(Past year pain reliever use disorder severe severity | past year use but not misuse of pain relievers and pain reliever use disorder)

Table E.153 Imputation Details for Past Year Hallucinogen Use Disorder: DSM-5

Edited Past Year Disorder	Restrictions Based on Past Year Use or Needle Use Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	74	1. SUD1 2. SUD2
Total cases requiring imputation	-	74	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year hallucinogen use disorder | past year use of hallucinogens)

2. SUD2 = $P(\text{No past year hallucinogen use disorder} \mid \text{past year use of hallucinogens})$

Table E.154 Imputation Details for Past Year Hallucinogen Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	1	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	0	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	8	1. SEV1 2. SEV2 3. SEV3
	No use disorder	67	N/A
Total cases requiring imputation	-	76	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = $P(\text{Past year hallucinogen use disorder mild severity} \mid \text{past year hallucinogen use disorder})$
2. SEV2 = $P(\text{Past year hallucinogen use disorder moderate severity} \mid \text{past year hallucinogen use disorder})$
3. SEV3 = $P(\text{Past year hallucinogen use disorder severe severity} \mid \text{past year hallucinogen use disorder})$

Table E.155 Imputation Details for Past Year Inhalant Use Disorder: DSM-5

Edited Past Year Disorder	Restrictions Based on Past Year Use or Needle Use Indicator	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	76	1. SUD1 2. SUD2
Total cases requiring imputation	-	76	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = $P(\text{Past year inhalant use disorder} \mid \text{past year use of inhalants})$
2. SUD2 = $P(\text{No past year inhalant use disorder} \mid \text{past year use of inhalants})$

Table E.156 Imputation Details for Past Year Inhalant Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	None	0	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	None	1	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	None	14	1. SEV1 2. SEV2 3. SEV3
	No use disorder	64	N/A
Total cases requiring imputation	-	79	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year inhalant use disorder mild severity | past year inhalant use disorder)
2. SEV2 = P(Past year inhalant use disorder moderate severity | past year inhalant use disorder)
3. SEV3 = P(Past year inhalant use disorder severe severity | past year inhalant use disorder)

Table E.157 Imputation Details for Past Year Tranquilizer Use Disorder: DSM-5

Edited Past Year Use Disorder	Imputation-Revised Past Year Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	Past year misuse	25	1. SUD1 2. SUD2
Completely missing	Past year use but not misuse	126	1. UNM1 2. UNM2
Total cases requiring imputation	-	151	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year tranquilizer use disorder | past year misuse of tranquilizers)
2. SUD2 = P(No past year tranquilizer use disorder | past year misuse of tranquilizers)
1. UNM1 = P(Past year tranquilizer use disorder | past year use but not misuse of tranquilizers)
2. UNM2 = P(No past year tranquilizer use disorder | past year use but not misuse of tranquilizers)

Table E.158 Imputation Details for Past Year Tranquilizer Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Imputation-Revised Past Year Use	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	Past year misuse	None	2	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	Past year misuse	None	1	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	Past year misuse	None	11	1. SEV1 2. SEV2 3. SEV3
		No disorder	15	N/A
Mild or moderate severity	Past year use but not misuse	None	0	1. UNMS1/(UNMS1+UNMS2) 2. UNMS2/(UNMS1+UNMS2)
Moderate or severe severity	Past year use but not misuse	None	2	1. UNMS2/(UNMS2+UNMS3) 2. UNMS3/(UNMS2+UNMS3)
Completely missing severity	Past year use but not misuse	None	12	1. UNMS1 2. UNMS2 3. UNMS3
		No disorder	115	N/A
Total cases requiring imputation	-	-	158	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year tranquilizer use disorder mild severity | past year misuse of tranquilizers and tranquilizer use disorder)
2. SEV2 = P(Past year tranquilizer use disorder moderate severity | past year misuse of tranquilizers and tranquilizer use disorder)
3. SEV3 = P(Past year tranquilizer use disorder severe severity | past year misuse of tranquilizers and tranquilizer use disorder)
1. UNMS1 = P(Past year tranquilizer use disorder mild severity | past year use but not misuse of tranquilizers and tranquilizer use disorder)
2. UNMS2 = P(Past year tranquilizer use disorder moderate severity | past year use but not misuse of tranquilizers and tranquilizer use disorder)
3. UNMS3 = P(Past year tranquilizer use disorder severe severity | past year use but not misuse of tranquilizers and tranquilizer use disorder)

Table E.159 Imputation Details for Past Year Sedative Use Disorder: DSM-5

Edited Past Year Use Disorder	Imputation-Revised Past Year Use	Number of Cases	Predictive Mean Vector ¹
Completely missing	Past year misuse	6	1. SUD1 2. SUD2
Completely missing	Past year use but not misuse	90	1. UNM1 2. UNM2
Total cases requiring imputation	-	96	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013).

¹ The predictive mean vector components are defined by the following:

1. SUD1 = P(Past year sedative use disorder | past year misuse of sedatives)
2. SUD2 = P(No past year sedative use disorder | past year misuse of sedatives)
1. UNM1 = P(Past year sedative use disorder | past year use but not misuse of sedatives)
2. UNM2 = P(No past year sedative use disorder | past year use but not misuse of sedatives)

Table E.160 Imputation Details for Past Year Sedative Use Disorder Severity: DSM-5

Edited Past Year Use Disorder Severity	Imputation-Revised Past Year Use	Restrictions Based on Imputation-Revised Substance Use Disorder	Number of Cases	Predictive Mean Vector ¹
Mild or moderate severity	Past year misuse	None	0	1. SEV1/(SEV1+SEV2) 2. SEV2/(SEV1+SEV2)
Moderate or severe severity	Past year misuse	None	0	1. SEV2/(SEV2+SEV3) 2. SEV3/(SEV2+SEV3)
Completely missing	Past year misuse	None	1	1. SEV1 2. SEV2 3. SEV3
		No disorder	5	N/A
Mild or moderate severity	Past year use but not misuse	None	1	1. UNMS1/(UNMS1+UNMS2) 2. UNMS2/(UNMS1+UNMS2)
Moderate or severe severity	Past year use but not misuse	None	1	1. UNMS2/(UNMS2+UNMS3) 2. UNMS3/(UNMS2+UNMS3)
Completely missing severity	Past year use but not misuse	None	8	1. UNMS1 2. UNMS2 3. UNMS3
		No disorder	83	N/A
Total cases requiring imputation	-	-	99	-

DSM-5 = 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2013); N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. SEV1 = P(Past year sedative use disorder mild severity | past year misuse of sedatives and sedative use disorder)
2. SEV2 = P(Past year sedative use disorder moderate severity | past year misuse of sedatives and sedative use disorder)
3. SEV3 = P(Past year sedative use disorder severe severity | past year misuse of sedatives and sedative use disorder)
1. UNMS1 = P(Past year sedative use disorder mild severity | past year use but not misuse of sedatives and sedative use disorder)
2. UNMS2 = P(Past year sedative use disorder moderate severity | past year use but not misuse of sedatives and sedative use disorder)
3. UNMS3 = P(Past year sedative use disorder severe severity | past year use but not misuse of sedatives and sedative use disorder)

E.3.2.5 Substance Use Treatment Variables

Table E.161 Imputation Details for Inpatient Substance Use Treatment in a Hospital

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	882	1. SUT1 2. SUT2
Total cases requiring imputation	882	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Inpatient treatment in a hospital in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No inpatient treatment in a hospital in the past year | lifetime drug or alcohol use)

Table E.162 Imputation Details for Inpatient Substance Use Treatment in a Residential Drug or Alcohol Rehab or Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	872	1. SUT1 2. SUT2
Total cases requiring imputation	872	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Inpatient treatment in a residential drug or alcohol rehab center in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No inpatient treatment in a residential drug or alcohol rehab center in the past year | lifetime drug or alcohol use)

Table E.163 Imputation Details for Inpatient Substance Use Treatment in a Residential Mental Health Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	870	1. SUT1 2. SUT2
Total cases requiring imputation	870	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Inpatient treatment in a residential mental health treatment center in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No inpatient treatment in a residential mental health treatment center in the past year | lifetime drug or alcohol use)

Table E.164 Imputation Details for Substance Use Treatment in Some Other Place as an Inpatient

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	870	1. SUT1 2. SUT2
Total cases requiring imputation	870	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Inpatient treatment in some other place in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No inpatient treatment in some other place in the past year | lifetime drug or alcohol use)

Table E.165 Imputation Details for Outpatient Substance Use Treatment at a Drug or Alcohol Rehab Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	871	1. SUT1 2. SUT2
Total cases requiring imputation	871	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Outpatient treatment at a rehab center in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No outpatient treatment at a rehab center in the past year | lifetime drug or alcohol use)

Table E.166 Imputation Details for Outpatient Substance Use Treatment at a Mental Health Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	879	1. SUT1 2. SUT2
Total cases requiring imputation	879	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Outpatient treatment at a mental health treatment center in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No outpatient treatment at a mental health treatment center in the past year | lifetime drug or alcohol use)

Table E.167 Imputation Details for Outpatient Substance Use Treatment at an Office of a Therapist, Psychologist, Psychiatrist, or Substance Use Treatment Professional

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	875	1. SUT1 2. SUT2
Total cases requiring imputation	875	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Outpatient treatment in a therapist's office in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No outpatient treatment in a therapist's office in the past year | lifetime drug or alcohol use)

Table E.168 Imputation Details for Outpatient Substance Use Treatment at a General Medical Clinic or Doctor's Office

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	879	1. SUT1 2. SUT2
Total cases requiring imputation	879	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Outpatient treatment at a doctor's office in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No outpatient treatment at a doctor's office in the past year | lifetime drug or alcohol use)

Table E.169 Imputation Details for Outpatient Substance Use Treatment at a Hospital

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	888	1. SUT1 2. SUT2
Total cases requiring imputation	888	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Outpatient treatment at a hospital in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No outpatient treatment at a hospital in the past year | lifetime drug or alcohol use)

Table E.170 Imputation Details for Outpatient Substance Use Treatment at a School Health or Counseling Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	876	1. SUT1 2. SUT2
Total cases requiring imputation	876	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Outpatient treatment at school in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No outpatient treatment at school in the past year | lifetime drug or alcohol use)

Table E.171 Imputation Details for Substance Use Treatment at Some Other Place as an Outpatient

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	892	1. SUT1 2. SUT2
Total cases requiring imputation	892	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Outpatient treatment in some other place in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No outpatient treatment in some other place in the past year | lifetime drug or alcohol use)

Table E.172 Imputation Details for Substance Use Treatment as Part of a Support Group

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	822	1. SUT1 2. SUT2
Total cases requiring imputation	822	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(treatment from a support group in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No treatment from a support group in the past year | lifetime drug or alcohol use)

Table E.173 Imputation Details for Substance Use Treatment in a Hospital Emergency Department

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	768	1. SUT1 2. SUT2
Total cases requiring imputation	768	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Emergency department treatment in the past year | lifetime drug or alcohol use)
2. SUT2 = P(No emergency department treatment in the past year | lifetime drug or alcohol use)

Table E.174 Imputation Details for Medication-Assisted Treatment for Alcohol Use

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	695	1. SUT1 2. SUT2
Total cases requiring imputation	695	-

¹ The predictive mean vector components are defined by the following:

1. SUT1 = P(Medication-assisted treatment for alcohol in the past year | lifetime alcohol use)
2. SUT2 = P(No medication-assisted treatment for alcohol in the past year | lifetime alcohol use)

Table E.175 Imputation Details for Medication-Assisted Treatment for Opioid Use

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	610	1. SUT1 2. SUT2
Total cases requiring imputation	610	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Medication-assisted treatment for opioids in the past year | lifetime heroin or prescription pain reliever use)
 2. SUT2 = P(No medication-assisted treatment for opioids in the past year | lifetime heroin or prescription pain reliever use)

Table E.176 Imputation Details for Use of Telehealth Substance Use Treatment

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	854	1. SUT1 2. SUT2
Total cases requiring imputation	854	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Telehealth treatment in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No telehealth treatment in the past year | lifetime drug or alcohol use)

Table E.177 Imputation Details for Substance Use Treatment in Prison, Jail, or Juvenile Detention Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	878	1. SUT1 2. SUT2
Total cases requiring imputation	878	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Treatment in prison in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No treatment in prison in the past year | lifetime drug or alcohol use)

Table E.178 Imputation Details for Use of Detoxification or Withdrawal Support Services

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,009	1. SUT1 2. SUT2
Total cases requiring imputation	1,009	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Detox in the past year | past year drug or alcohol use)
 2. SUT2 = P(No detox in the past year | past year drug or alcohol use)

Table E.179 Imputation Details for Substance Use Treatment from a Peer Support Specialist or Recovery Coach

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	873	1. SUT1 2. SUT2
Total cases requiring imputation	873	-

- ¹ The predictive mean vector components are defined by the following:
1. SUT1 = P(Treatment from a peer support specialist in the past year | lifetime drug or alcohol use)
 2. SUT2 = P(No treatment from a peer support specialist in the past year | lifetime drug or alcohol use)

E.4 Mental Health Variables

E.4.1 Adult Mental Health Variables

Adult mental health variables were imputed with modPMN (see Chapter 3 for details). The tables in this section present information on the imputation procedures applied under the modPMN method. For continuous variables, information on the missingness patterns, constraints, and predictive mean vectors used for hot-deck imputation is presented. For categorical variables, information on the edited values, restrictions from previously imputed variables, and predictive mean vectors used for CSI is presented.

Table E.180 Likeness Constraints for Kessler-6 (K6) Psychological Distress Scale

Constraint #	Likeness Constraint
LikC1	Donor's value must match recipient's value for each nonmissing item in the K6 scale
LikC2	Donor's total K6 score must be within the range defined by the recipient's minimum and maximum possible K6 score.
	Recipient's minimum score is equal to the sum of all nonmissing items in the K6 scale
	Recipient's maximum score is equal to the minimum score plus four times the number of missing items in the K6 scale
LikC3	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC4	Donor's age is within 5 years of recipient's age
LikC5	State rank of donor = state rank of recipient
LikC6	Donor and recipient are in the same RP bin

Note: The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

Table E.181 Constraints and Portion of the Predictive Mean Vector for Past Month Kessler-6 (K6) Psychological Distress Scale

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 3
Cycle 1									
1	Missing one or more of the items in the K6 scale	1,505	None	1. PK6PM	1-6: 40	1-6: 133	1-6: 201	1-6: 325	1-6: 628
					1-4,6: 4	1-4,6: 9	1-4,6: 6	1-4,6: 8	1-4,6: 11
					1-3,6: 2	1-3,6: 5	1-3,6: 10	1-3,6: 11	1-3,6: 11
					1,2,6: 4	1,2,6: 18	1,2,6: 19	1,2,6: 21	1,2,6: 25
						1,6: 0	1,6: 0	1,6: 0	1,6: 0
		2,6: 2	2,6: 3	2,6: 4	2,6: 5				
Cycle 2									
1	Missing one or more of the items in the K6 scale	1,505	None	1. PK6PM	1-6: 40	1-6: 129	1-6: 203	1-6: 322	1-6: 632
					1-4,6: 4	1-4,6: 9	1-4,6: 12	1-4,6: 7	1-4,6: 15
					1-3,6: 4	1-3,6: 8	1-3,6: 11	1-3,6: 8	1-3,6: 10
					1,2,6: 4	1,2,6: 17	1,2,6: 12	1,2,6: 18	1,2,6: 26
						1,6: 0	1,6: 0	1,6: 0	1,6: 0
		2,6: 2	2,6: 3	2,6: 4	2,6: 5				

Note: The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

¹ The predictive mean vector components are defined by the following:

1. PK6PM = Predicted K6 total score in past month

Table E.182 Constraints and Portion of the Predictive Mean Vector for Worst Month in Past Year Kessler-6 (K6) Psychological Distress Scale

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 3
Cycle 1									
1	Missing one or more of the items in the K6 scale	550	None	1. PK6PY	1-6: 25	1-6: 51	1-6: 58	1-6: 114	1-6: 222
					1-4,6: 2	1-4,6: 4	1-4,6: 6	1-4,6: 3	1-4,6: 9
					1-3,6: 1	1-3,6: 1	1-3,6: 6	1-3,6: 7	1-3,6: 6
					1,2,6: 5	1,2,6: 5	1,2,6: 2	1,2,6: 6	1,2,6: 7
					1,6: 0	1,6: 0	1,6: 0	1,6: 0	1,6: 0
					2,6: 1	2,6: 5	2,6: 1	2,6: 1	2,6: 2
Cycle 2									
1	Missing one or more of the items in the K6 scale	513	None	1. PK6PY	1-6: 22	1-6: 36	1-6: 54	1-6: 90	1-6: 230
					1-4,6: 2	1-4,6: 3	1-4,6: 3	1-4,6: 4	1-4,6: 6
					1-3,6: 1	1-3,6: 5	1-3,6: 3	1-3,6: 9	1-3,6: 7
					1,2,6: 2	1,2,6: 4	1,2,6: 8	1,2,6: 10	1,2,6: 7
						1,6: 0	1,6: 0		1,6: 0
						2,6: 3	2,6: 3		2,6: 1

Note: The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

¹ The predictive mean vector components are defined by the following:

1. PK6PY = Predicted K6 total score in worst month in past year

Table E.183 Imputation Details for Kessler-6 (K6) Psychological Distress Scale, Month in Past Year Worse than Past 30 Days

Edited Value	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,340	1. PW1 2. PW2
Total cases requiring imputation	1,340	-

Note: The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

¹ The predictive mean vector components are defined by the following:

1. PW1 = P(Month in past year worse than past 30 days)
2. PW2 = P(No month in past year worse than past 30 days)

Table E.184 Logical Constraints for WHODAS

Constraint #	Logical Constraint
LogC1	If the recipient reported not participating in an activity and is missing the follow-up question about whether emotional problems kept the recipient from participating, then the donor must not have a legitimate skip code for the follow-up question
LogC2	If recipient reported a level of difficulty taking care of responsibilities at work or school (IMPRES=1,2,3,4) but did not answer the question pertaining to getting work done as quickly as needed (IMPWORK=missing), then the donor must not have a value indicating the donor did not go to work or school (IMPRES ≠ 5) If the recipient reported a level of difficulty getting work done as quickly as needed (IMPWORK=1,2,3,4) but did not answer the question for whether the recipient had difficulty taking care of responsibilities at work or school (IMPRES=missing), then the donor must not have a value indicating the donor did not go to work or school (IMPRES ≠ 5)

Note: The World Health Organization Disability Assessment Schedule (WHODAS) is a scale used to measure functional impairment due to disturbances in social adjustment and behavior (Novak et al., 2010; Rehm et al., 1999).

Note: Respondents reported experiencing at least one of the Kessler-6 (K6) symptoms at least a “little of the time.” The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

Table E.185 Likeness Constraints for WHODAS

Constraint #	Logical Constraint
LikC1	Donor's value must match recipient's value for each nonmissing item in the WHODAS scale
LikC2	Donor's total WHODAS score must be within the range defined by the recipient's minimum and maximum possible WHODAS score Recipient's minimum score is equal to the sum of all nonmissing scored items in the WHODAS scale Recipient's maximum score is equal to the minimum score plus three times the number of missing items in the WHODAS scale
LikC3	Donor's alternative WHODAS score must be within the range defined by the recipient's minimum and maximum possible alternative WHODAS score Recipient's minimum alternative WHODAS score is equal to the number of daily activities for which the respondent reported having moderate or severe difficulty performing Recipient's maximum alternative WHODAS score is equal to the minimum score plus the number of activities with missing data
LikC4	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC5	Donor's age is within 5 years of recipient's age
LikC6	State rank of donor = state rank of recipient
LikC7	Donor and recipient are in the same response propensity bin

Note: The World Health Organization Disability Assessment Schedule (WHODAS) is a scale used to measure functional impairment due to disturbances in social adjustment and behavior (Novak et al., 2010; Rehm et al., 1999).

Note: Respondents reported experiencing at least one of the Kessler-6 (K6) symptoms at least a "little of the time." The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

Table E.186 Constraints and Portion of the Predictive Mean Vector for WHODAS

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Response Propensity Bin				
					Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
1	Missing one or more of the items in the WHODAS scale	1,457	1,2	1. PWSC 2. PWDA	1-7: 27	1-7: 83	1-7: 152	1-7: 203	1-7: 460
					1-5,7: 6	1-5,7: 13	1-5,7: 26	1-5,7: 38	1-5,7: 63
					1-4,7: 7	1-4,7: 11	1-4,7: 18	1-4,7: 30	1-4,7: 39
					1-3,7: 20	1-3,7: 31	1-3,7: 33	1-3,7: 33	1-3,7: 66
					1,3,7: 0	1,3,7: 0	1,3,7: 0	1,3,7: 0	1,3,7: 0
					1,2,7: 0	1,2,7: 0	1,2,7: 0	1,2,7: 0	1,2,7: 0
					2,3,7: 7	2,3,7: 21	2,3,7: 15	2,3,7: 25	2,3,7: 30

Note: The World Health Organization Disability Assessment Schedule (WHODAS) is a scale used to measure functional impairment due to disturbances in social adjustment and behavior (Novak et al., 2010; Rehm et al., 1999).

Note: Respondents reported experiencing at least one of the Kessler-6 (K6) symptoms at least a “little of the time.” The K6 screening instrument is used to measure nonspecific psychological distress (Furukawa et al., 2003; Kessler et al., 2003).

¹ The predictive mean vector components are defined by the following:

1. PWSC = Predicted WHODAS total score
2. PWDA = Predicted alternative WHODAS score

Table E.187 Imputation Details for Lifetime MDE

Edited Lifetime MDE	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,525	1. MDE1 2. MDE2
Total cases requiring imputation	1,525	-

MDE = major depressive episode; N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. MDE1= P(MDE in lifetime)
2. MDE2= P(No MDE in lifetime)

Table E.188 Imputation Details for Past Year MDE

Edited Past Year MDE	Restrictions Based on Imputation-Revised Lifetime MDE	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	433	1. MDE1 2. MDE2
	No past year MDE	1,205	N/A
Total cases requiring imputation	-	1,638	-

MDE = major depressive episode; N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. MDE1= P(MDE in past year | MDE in lifetime)
2. MDE2= P(No MDE in past year | MDE in lifetime)

Table E.189 Imputation Details for Past Year MDE with Impairment

Edited Past Year MDE with Impairment	Restrictions Based on Imputation-Revised Past Year MDE	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	256	1. MDE1 2. MDE2
	No past year MDE with Impairment	1,412	N/A
Total cases requiring imputation	-	1,668	-

MDE = major depressive episode; N/A = not applicable.

¹ The predictive mean vector components are defined by the following:

1. MDE1= P(MDE in past year with impairment | past year MDE)
2. MDE2= P(No MDE in past year with impairment | past year MDE)

Table E.190 Imputation Details for Suicidal Thoughts

Edited Suicidal Thoughts	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,378	1. SUI1 2. SUI2
Total cases requiring imputation	1,378	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal thoughts in past 12 months)
2. SUI2= P(No suicidal thoughts in past 12 months)

Table E.191 Imputation Details for Suicidal Plans

Edited Suicidal Plans	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,304	1. SUI1 2. SUI2
Total cases requiring imputation	1,304	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal plans in past 12 months)
2. SUI2= P(No suicidal plans in past 12 months)

Table E.192 Imputation Details for Suicidal Attempts

Edited Suicidal Attempts	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,285	1. SUI1 2. SUI2
Total cases requiring imputation	1,285	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal attempts in past 12 months)
2. SUI2= P(No suicidal attempts in past 12 months)

Table E.193 Imputation Details for Suicidal Thoughts due to Coronavirus Disease 2019 (COVID-19)

Edited Suicidal Thoughts due to COVID-19	Restrictions Based on Imputation-Revised Suicidal Thoughts	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	107	1. SUI1 2. SUI2
	No suicidal thoughts in past 12 months due to COVID-19	1,291	N/A
Total cases requiring imputation	-	1,398	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal thoughts due to COVID-19 in past 12 months | suicidal thoughts in past 12 months)
2. SUI2= P(No suicidal thoughts due to COVID-19 in past 12 months | suicidal thoughts in past 12 months)

Table E.194 Imputation Details for Suicidal Plans due to Coronavirus Disease 2019 (COVID-19)

Edited Suicidal Plans due to COVID-19	Restrictions Based on Imputation-Revised Suicidal Plans	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	27	1. SUI1 2. SUI2
	No suicidal plans in past 12 months due to COVID-19	1,284	N/A
Total cases requiring imputation	-	1,311	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal plans due to COVID-19 in past 12 months | suicidal plans in past 12 months)
2. SUI2= P(No suicidal plans due to COVID-19 in past 12 months | suicidal plans in past 12 months)

Table E.195 Imputation Details for Suicidal Attempts due to Coronavirus Disease 2019 (COVID-19)

Edited Suicidal Attempts due to COVID-19	Restrictions Based on Imputation-Revised Suicidal Attempts	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	14	1. SUI1 2. SUI2
	No suicidal attempts in past 12 months due to COVID-19	1,274	N/A
Total cases requiring imputation	-	1,288	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Suicidal attempts due to COVID-19 in past 12 months | suicidal attempts in past 12 months)
2. SUI2= P(No suicidal attempts due to COVID-19 in past 12 months | suicidal attempts in past 12 months)

Table E.196 Imputation Details for Receive Medical Attention for Suicidal Attempt

Edited Medical Attention for Suicidal Attempt	Restrictions Based on Imputation-Revised Suicidal Attempts	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	11	1. SUI1 2. SUI2
	No suicidal attempts in past 12 months	1,274	N/A
Total cases requiring imputation	-	1,285	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Received medical attention for suicidal attempt in past 12 months | suicidal attempt(s) in past 12 months)
2. SUI2= P(No received medical attention for suicidal attempt in past 12 months | suicidal attempt(s) in past 12 months)

Table E.197 Imputation Details for Overnight or Longer Hospital Stay for Suicidal Attempt

Edited Hospital Stay for Suicidal Attempt	Restrictions Based on Imputation-Revised Medical Attention for Suicidal Attempt	Number of Cases	Predictive Mean Vector ¹
Completely missing	None	8	1. SUI1 2. SUI2
	No medical attention for suicidal attempt in past 12 months	1,277	N/A
Total cases requiring imputation	-	1,285	-

¹ The predictive mean vector components are defined by the following:

1. SUI1= P(Overnight or longer hospital stay for suicidal attempt in past 12 months | medical attention for suicidal attempt in past 12 months)
2. SUI2= P(No overnight or longer hospital stay for suicidal attempt in past 12 months | medical attention for suicidal attempt in past 12 months)

E.4.2 Mental Health Treatment Variables

Mental health treatment variables were imputed with modPMN (see Chapter 3 for details). All of these variables are dichotomous, and none have any restrictions based on previously imputed variables. The tables in this section present the predictive mean vectors used for CSI.

Table E.198 Imputation Details for Inpatient Mental Health Treatment in a Hospital

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,291	1. MHT1 2. MHT2
Total cases requiring imputation	1,291	-

¹ The predictive mean vector components are defined by the following:

1. MHT1 = P(Inpatient treatment in a hospital in the past year)
2. MHT2 = P(No inpatient treatment in a hospital in the past year)

Table E.199 Imputation Details for Inpatient Mental Health Treatment in a Residential Mental Health Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,276	1. MHT1 2. MHT2
Total cases requiring imputation	1,276	-

¹ The predictive mean vector components are defined by the following:

1. MHT1 = P(Inpatient treatment in a residential mental health treatment center in the past year)
2. MHT2 = P(No inpatient treatment in a residential mental health treatment center in the past year)

Table E.200 Imputation Details for Inpatient Mental Health Treatment in a Residential Drug or Alcohol Rehab or Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,246	1. MHT1 2. MHT2
Total cases requiring imputation	1,246	-

¹ The predictive mean vector components are defined by the following:

1. MHT1 = P(Inpatient treatment in a residential drug or alcohol rehab center in the past year)
2. MHT2 = P(No inpatient treatment in a residential drug or alcohol rehab center in the past year)

Table E.201 Imputation Details for Mental Health Treatment in Some Other Place as an Inpatient

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,271	1. MHT1 2. MHT2
Total cases requiring imputation	1,271	-

¹ The predictive mean vector components are defined by the following:

1. MHT1 = P(Inpatient treatment in some other place in the past year)
2. MHT2 = P(No inpatient treatment in some other place in the past year)

Table E.202 Imputation Details for Outpatient Mental Health Treatment at a Mental Health Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,329	1. MHT1 2. MHT2
Total cases requiring imputation	1,329	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment at a mental health treatment center in the past year)
 2. MHT2 = P(No outpatient treatment at a mental health treatment center in the past year)

Table E.203 Imputation Details for Outpatient Mental Health Treatment at a Drug or Alcohol Rehab Treatment Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,279	1. MHT1 2. MHT2
Total cases requiring imputation	1,279	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment at a rehab center in the past year)
 2. MHT2 = P(No outpatient treatment at a rehab center in the past year)

Table E.204 Imputation Details for Outpatient Mental Health Treatment at an Office of a Therapist, Psychologist, Psychiatrist, or Mental Health Treatment Professional

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,309	1. MHT1 2. MHT2
Total cases requiring imputation	1,309	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment in a therapist's office in the past year)
 2. MHT2 = P(No outpatient treatment in a therapist's office in the past year)

Table E.205 Imputation Details for Outpatient Mental Health Treatment at a General Medical Clinic or Doctor's Office

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,330	1. MHT1 2. MHT2
Total cases requiring imputation	1,330	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment at a doctor's office in the past year)
 2. MHT2 = P(No outpatient treatment at a doctor's office in the past year)

Table E.206 Imputation Details for Outpatient Mental Health Treatment at a Hospital

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,333	1. MHT1 2. MHT2
Total cases requiring imputation	1,333	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment at a hospital in the past year)
 2. MHT2 = P(No outpatient treatment at a hospital in the past year)

Table E.207 Imputation Details for Outpatient Mental Health Treatment at a School Health or Counseling Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,325	1. MHT1 2. MHT2
Total cases requiring imputation	1,325	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment at school in the past year)
 2. MHT2 = P(No outpatient treatment at school in the past year)

Table E.208 Imputation Details for Mental Health Treatment at Some Other Place as an Outpatient

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,373	1. MHT1 2. MHT2
Total cases requiring imputation	1,373	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Outpatient treatment in some other place in the past year)
 2. MHT2 = P(No outpatient treatment in some other place in the past year)

Table E.209 Imputation Details for Mental Health Treatment as Part of a Support Group

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,220	1. MHT1 2. MHT2
Total cases requiring imputation	1,220	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(treatment from a support group in the past year)
 2. MHT2 = P(No treatment from a support group in the past year)

Table E.210 Imputation Details for Mental Health Treatment in a Hospital Emergency Department

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,142	1. MHT1 2. MHT2
Total cases requiring imputation	1,142	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Emergency department treatment in the past year)
 2. MHT2 = P(No emergency department treatment in the past year)

Table E.211 Imputation Details for Use of Prescription Medication for Mental Health Treatment

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,185	1. MHT1 2. MHT2
Total cases requiring imputation	1,185	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Past year use of prescription medication for mental health treatment)
 2. MHT2 = P(No past year use of prescription medication for mental health treatment)

Table E.212 Imputation Details for Use of Telehealth Mental Health Treatment

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,198	1. MHT1 2. MHT2
Total cases requiring imputation	1,198	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Telehealth treatment in the past year)
 2. MHT2 = P(No telehealth treatment in the past year)

Table E.213 Imputation Details for Mental Health Treatment in Prison, Jail, or Juvenile Detention Center

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,180	1. MHT1 2. MHT2
Total cases requiring imputation	1,180	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Treatment in prison in the past year)
 2. MHT2 = P(No treatment in prison in the past year)

Table E.214 Imputation Details for Mental Health Treatment from a Peer Support Specialist or Recovery Coach

Edited Treatment in Past Year	Number of Cases	Predictive Mean Vector ¹
Completely missing	1,273	1. MHT1 2. MHT2
Total cases requiring imputation	1,273	-

- ¹ The predictive mean vector components are defined by the following:
1. MHT1 = P(Treatment from a peer support specialist in the past year)
 2. MHT2 = P(No treatment from a peer support specialist in the past year)

E.5 Household Composition (Roster) Variables

The tables in this section present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the six household composition (roster) variables: number of rostered people, number of children younger than 18, number of people aged 65 or older, indicator of whether the respondent has family members in household, number of respondent's family members in the household (including foster relationships), and number of respondent's family members younger than 18 in the household (including foster relationships).

Table E.215 Logical Constraints for Number of Rostered People

Constraint #	Logical Constraint
LogC1	Lower bound was restricted based on nonmissing roster members

Table E.216 Likeness Constraints for Number of Rostered People

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.217 Constraints and Portion of the Predictive Mean Vector for Number of Rostered People

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	1,545	1	1. C1	1: 39	1: 428	1: 941	1: 136 None: 1

¹ The predictive mean vector components are defined by the following:

1. C1 = Predicted count from Poisson regression model

Table E.218 Logical Constraints for Number of Children Younger than 18

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.219 Likeness Constraints for Number of Children Younger than 18

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.220 Constraints and Portion of the Predictive Mean Vector for Number of Children Younger than 18

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	1,464	1	1. C1	1,2: 77 2: 2	1,2: 399 2: 2	1,2: 882 2: 4	1,2: 98

¹ The predictive mean vector components are defined by the following:

1. C1 = Predicted count from Poisson regression model

Table E.221 Logical Constraints for Number of People Aged 65 or Older

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE, IRKID17, and nonmissing ages in the roster

Table E.222 Likeness Constraints for Number of People Aged 65 or Older

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.223 Constraints and Portion of the Predictive Mean Vector for Number of People Aged 65 or Older

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	1,297	1	1. C1	1,2: 61 1: 1	1,2: 373	1,2: 763 1: 2	1,2: 97

¹ The predictive mean vector components are defined by the following:

1. C1 = Predicted count from Poisson regression model

Table E.224 Likeness Constraints for Indicator of Whether the Respondent Has Family Members in Household

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	Donor's predicted means must be within 20 percent of recipient's predicted means
LikC3	IRKID17 of donor = IRKID17 of recipient
LikC4	If recipient was married, then donor was married; otherwise, if recipient was not currently married, then donor was not currently married

Table E.225 Constraints and Portion of the Predictive Mean Vector for Indicator of Whether the Respondent Has Family Members in Household

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	1,229	None	1. F1	1-4: 23	1-4: 340 2-4: 2	1-4: 770 2-4: 0 2,4: 1	1-4: 91 2-4: 2

¹ The predictive mean vector components are defined by the following:

1. F1 = P(No other family members in the household)

Table E.226 Logical Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.227 Likeness Constraints for Number of Respondent's Family Members in Household (Including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.228 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (Including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	1,427	1	1. C1	1-3: 69 2,3: 0 3: 1	1-3: 387 2,3: 2 3: 0 None: 0 None*: 1	1-3: 867 2,3: 2	1-3: 98

*A donor could not be found for this case. The imputed value was randomly assigned within bounds.

¹ The predictive mean vector components are defined by the following:

1. C1 = Predicted count from Poisson regression model

Table E.229 Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFMLYSZ and nonmissing ages in the roster

Table E.230 Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient

Table E.231 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (Including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	675	1	1. C1	1-4: 57 2-4: 0 3,4: 0 4: 1	1-4: 127 2-4: 1 3,4: 0 4: 1 None: 1	1-4: 476 2-4: 5	1-4: 6

¹ The predictive mean vector components are defined by the following:

1. C1 = Predicted count from Poisson regression model

E.6 Income Variables

The tables in this section present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the income variables.

E.6.1 Binary Variable Phase

Five of the binary income variables were directly related to a respondent's socioeconomic status. Hence, if a recipient required imputation for one or more of these five variables (i.e., welfare payments, welfare services, Supplemental Nutrition Assistance Program [SNAP] benefits, binary income, and months on welfare), but had information on at least one of these variables, then the donors were restricted so that donors and recipients had the same values for these nonmissing variables. These five variables are referred to as "welfare-correlated variables."

There were a large number of missingness patterns for the source-of-income variables because they are imputed simultaneously in a set. A respondent could be missing any combination of the nine source-of-income variables. Because the constraints and predictive mean vectors can be described easily, only one row in [Table E.234](#) is used to summarize the many multiple variable missingness patterns.

Table E.232 Logical Constraints for Source of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient ¹
LogC2	Recipient is missing months on welfare. If recipient is also known to have received either welfare payments or welfare services, then donor must also have received welfare payments or welfare services. (This prevents donor from giving a skip code to a recipient who should have a value from 1 to 12.)

¹ This is only a logical constraint when family binary total income is missing.

Table E.233 Likeness Constraints for Source of Income

Constraint #	Likeness Constraint
LikC1	Age of donor = age of recipient
LikC2	If any of the welfare-correlated edited binary income variables (welfare payments, welfare services, SNAP benefits, binary total income, and months on welfare) were missing, then donor must match recipient on all nonmissing welfare-correlated edited binary income variables
LikC3	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing family variables
LikC4	Donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing)
LikC5	If recipient is missing social security, then donor must match recipient with respect to whether there are adults aged 65 or older in the household
LikC6	If recipient is missing welfare payments and/or welfare services, then donor must match recipient with respect to whether there are children younger than 18 in the household
LikC7	Age of donor must be within 5 years of age of recipient
LikC8	If recipient is missing binary income at the family level but not at the personal level, then donor must match recipient with respect to binary income at the personal level
LikC9	If recipient is not missing binary income at the personal level, then donor must match recipient with respect to binary income at the personal level

SNAP = Supplemental Nutrition Assistance Program.

Table E.234 Constraints and Portion of the Predictive Mean Vector for Source of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing welfare months, receiving family payment and/or family service	86	1,2	1. WMS, and probabilities associated with other missing elements	1-6: 10 1,2,4-6: 21	1-6: 4 1,2,4-6: 15 2,4-7: 2	1-6: 3 1,2,4-6: 22 2,4-7: 2 2,4-6: 1	1-6: 0 1,2,4-6: 2 2,4-7: 4
2	Missing welfare months, not receiving welfare payments, missing welfare services	109	1,2	1. SVC*WMS, SVC, and probabilities associated with other missing elements	1-6: 10 1,2,4-6: 20	1-6: 6 1,2,4-6: 41	1-6: 2 1,2,4-6: 26	1-6: 0 1,2,4-6: 4
3	Missing welfare months, missing welfare payments, not receiving welfare services	306	1,2	1. PMT*WMS, PMT, and probabilities associated with other missing elements	1-6: 30 1,2,4-6: 40	1-6: 31 1,2,4-6: 101	1-6: 21 1,2,4-6: 55	1-6: 6 1,2,4-6: 22
4	Missing welfare months, missing welfare payments, missing welfare services	1,942	1,2	1. [1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements	1-6: 10 1,2,4-6: 133	1-6: 7 1,2,4-6: 529 2,4-7: 1	1-6: 14 1,2,4-6: 1,084 2,4-7: 1	1-6: 4 1,2,4-6: 159
5	All other missingness patterns with number of months on welfare nonmissing ²	2,591	1	1. Probabilities associated with missing elements	1,2,3,5,6: 533 1,2,5,6: 36 2,5-7: 1	1,2,3,5,6: 842 1,2,5,6: 101	1,2,3,5,6: 664 1,2,5,6: 202 2,5-7: 3 2,5,6: 1	1,2,3,5,6: 164 1,2,5,6: 41 2,5-7: 3

¹ The predictive mean vector components are defined by the following:

1. PMT = P(Family in household received income from welfare payments)

2. SVC = P(Family in household received income from other welfare services)

3. WMS = P(Family in household received any welfare on a given month in the past year | family received any welfare in the past year)

² There were many other missingness patterns for source of income because a respondent could be missing any combination of the nine source-of-income variables. Because the constraints and predictive mean vectors can be described easily, all of these other missingness patterns are represented by a single row in the table.

E.6.2 Specific Category Phase

Table E.235 Logical Constraints for Finer Categories of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient
LogC2	Family income (FMINC3) of donor must be greater than personal income (PRINC3) of recipient
LogC3	Personal income (PRINC3) of donor must be less than family income (FMINC3) of recipient

Table E.236 Likeness Constraints for Finer Categories of Income

Constraint #	Likeness Constraint
LikC1	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC2	PRINC3 of donor = PRINC3 of recipient
LikC3	FMINC3 of donor = FMINC3 of recipient
LikC4	IRPINC1 of donor = IRPINC1 of recipient
LikC5	IRFINC1 of donor = IRFINC1 of recipient

Table E.237 Constraints and Portion of the Predictive Mean Vector for Finer Categories of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing personal income only	1,040	13	1. $X\beta$	1,3-5: 108	1,3-5: 287	1,3-5: 472	1,3-5: 172 3-5: 1
2	Missing family income only	3,494	1,2	1. $X\beta$	1,2,4,5: 750 2,4,5: 11	1,2,4,5: 1,359 2,4,5: 8	1,2,4,5: 1,166 2,4,5: 3	1,2,4,5: 187 2,4,5: 10
3	Completely missing	2,615	1	1. $X\beta$	1,4,5: 228	1,4,5: 665	1,4,5: 1,405	1,4,5: 317

¹ For more details, see Section 12.3.2.2.

E.7 Health Insurance Variables

The tables in this section present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the health insurance variables. Prior to 2013, two methods were presented for health insurance variables: the "constituent variables method" and the "old method." See Section 13.1 for an explanation of the "constituent variables method" and why the "old method" is no longer presented.

In several instances, variable names are used without description for the purposes of brevity (see Chapter 13 for details). For the health insurance imputations, matches between donors and recipients were attempted on the nonmissing values of the variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. These variables are the edited indicators of whether the respondent received health insurance from Medicaid/state health insurance programs for children, Medicare, CHAMPUS, or private health insurance, respectively. These were the base variables

used in the creation of the imputation-revised variables (IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT). In addition to the edited health insurance variables, other variables, which were used as likeness constraints, are identified in the tables only by their variable names. These include SERVICE (an indicator of whether the respondent had ever been in the military), GOVTPROG (an indicator of whether the respondent's family participated in government public assistance programs), INCOME (a four-level categorical family-income variable, with levels of less than \$20,000, \$20,000 to less than \$50,000, \$50,000 to less than \$75,000, and \$75,000 or more), IRFAMIN1 (a two-level family income variable, with levels of less than \$20,000 and \$20,000 or more), and IRFAMSOC (an indicator of whether the respondent's family in the household received income from social security).

For the multivariate predictive mean neighborhood (MPMN) method, the likeness constraints, which were applied to the variables, differed between missingness patterns, and sometimes the constraints differed between age groups within the same missingness pattern.

Table E.238 Likeness Constraints for Health Insurance (MPMN)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing insurance variables (CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN)
LikC2	GOVTPROG of donor = GOVTPROG of recipient
LikC3	IRFAMSOC of donor = IRFAMSOC of recipient
LikC4	SERVICE of donor = SERVICE of recipient (if nonmissing)
LikC5	INCOME of donor = INCOME of recipient
LikC6	IRFAMIN1 of donor = IRFAMIN1 of recipient
LikC7	Donor must match recipient for all nonmissing insurance variables
LikC8	If recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), and recipient has no job due to disability (JBSTATR = 14), then donor must have no job due to disability recipient has JBSTATR not equal to 14, then donor must have JBSTATR not equal to 14

MPMN = multivariate predictive mean neighborhood.

Table E.239 Constraints and Portion of the Predictive Mean Vector for Health Insurance (MPMN)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Only missing CAIDCHIP	306	None	1. HI1	1,2,7: 64 1,2: 1	1,2,7: 130 1,2: 4	1,2,7: 66 1,2: 1	1,2,7: 40
2	Only missing MEDICARE	123	None	1. HI2	1,3,7: 35 1,3: 1	1,3,7,8: 43	1,3,7,8: 36	1,3,7: 8
3	Missing CAIDCHIP and MEDICARE	72	None	1. HI1 2. HI2	1-3,7: 13 1-3: 4 1: 2 None: 2	1,2,7,8: 22 1,2,8: 4 1,8: 1 8: 3	1,2,7,8: 16 1,2,8: 2 1,8: 1	1,2,7: 1 1,2: 1
4	Only missing CHAMPUS	48	None	1. HI3	1,4,7: 19	1,4,7: 11	1,4,7: 14	1,4,7: 4
5	Missing CAIDCHIP and CHAMPUS	12	None	1. HI1 2. HI3	1,2,4,7: 4 1,2,4: 1	1,2,4,7: 5	1,2,4,7: 1 1,2,4: 1	No Cases
6	Missing MEDICARE and CHAMPUS	6	None	1. HI2 2. HI3	1,3,4,7: 3	1,4,7,8: 2	1,4,7,8: 0 1,4,8: 1	No Cases
7	Missing CAIDCHIP, MEDICARE, and CHAMPUS	12	None	1. HI1 2. HI2 3. HI3	1-4,7: 5 1-4: 0 1,4: 0 4: 2	1,2,4,7,8: 1 1,2,4,8: 1 1,4,8: 0 4,8: 3	No Cases	No Cases
8	Only missing PRVHLTIN	238	None	1. HI4	1,5,7: 56	1,5,7: 114 1,5: 4	1,5,7: 43 1,5: 1	1,6,7: 20
9	Missing CAIDCHIP and PRVHLTIN	41	None	1. HI1 2. HI4	1,2,5,7: 6 1,2,5: 1 1,6: 1 1: 0 None: 1	1,2,5,7: 20 1,2,5: 1 1,6: 2 1: 0 None: 2	1,2,5,7: 3	1,2,6,7: 4
10	Missing MEDICARE and PRVHLTIN	16	None	1. HI2 2. HI4	1,3,5,7: 8 1,3,5: 0 1,6: 1	1,5,7,8: 4	1,5,7,8: 1 1,5,8: 0 1,6,8: 0 1,8: 0 8: 2	No Cases

Table E.239 Constraints and Portion of the Predictive Mean Vector for Health Insurance (MPMN) (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
11	Missing CAIDCHIP, MEDICARE, and PRVHLTIN	34	None	1. HI1 2. HI2 3. HI4	1-3,5,7: 3	1,2,5,7,8: 1	1,2,5,7,8: 0	1,2,6,7: 0
					1-3,5: 0	1,2,5,8: 0	1,2,5,8: 0	1,2,6: 0
					1,6: 0	1,6,8: 0	1,6,8: 0	1: 0
					1: 0	1,8: 1	1,8: 0	None: 1
12	Missing CHAMPUS and PRVHLTIN	23	None	1. HI3 2. HI4	None: 14	8: 11	8: 3	
					1,4,5,7: 6	1,4,5,7: 6	1,4,5,7: 4	1,4,6,7: 1
						1,4,5: 1	1,4,5: 0	1,4,6: 0
						1,4,6: 1	1,4,6: 1	1,4: 0
13	Missing CAIDCHIP, CHAMPUS, and PRVHLTIN	43	None	1. HI1 2. HI3 3. HI4		1,4: 0	1,4: 1	4: 1
						4: 1		
					1,2,4,5,7: 8	1,2,4,5,7: 4	1,2,4,5,7: 0	No Cases
					1,2,4,5: 0	1,2,4,5: 0	1,2,4,5: 2	
14	Missing MEDICARE, CHAMPUS, and PRVHLTIN	4	None	1. HI2 2. HI3 3. HI4	1,4,6: 1	1,4,6: 3	1,4,6: 1	
					1,4: 0	1,4: 0	1,4: 1	
					4: 7	4: 8	4: 8	
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	1,713	None	1. HI1 2. HI2 3. HI3 4. HI4	1,3-5,7: 0	1,4,5,7,8: 0	No Cases	No Cases
					1,3-5: 0	1,4,5,8: 0		
					1,4,6: 0	1,4,6,8: 0		
					1,4: 0	1,4,8: 0		
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	1,713	None	1. HI1 2. HI2 3. HI3 4. HI4	4: 3	4,8: 1		
					1-5: 14	1,2,4,5,8: 39	1,2,4,5,8: 118	1,2,4,6: 44
					1,4,6: 0	1,4,6,8: 2	1,4,6,8: 7	1,4: 0
					1,4: 0	1,4,8: 2	1,4,8: 3	4: 111
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	1,713	None	1. HI1 2. HI2 3. HI3 4. HI4	4: 65	4,8: 415	4,8: 893	

MPMN = multivariate predictive mean neighborhood.

¹ The predictive mean vector components are defined by the following:

1. HI1 = P(Respondent has CAIDCHIP)
2. HI2 = P(Respondent has Medicare)
3. HI3 = P(Respondent has CHAMPUS)
4. HI4 = P(Respondent has private health insurance)

Table E.240 Constraints for Health Insurance (UPMN), Any Other Health Insurance

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is between 26 and 64 years old, then donor must also be between 26 and 64 years old; if recipient is aged 65 years or older, then donor must also be aged 65 years or older

UPMN = univariate predictive mean neighborhood.

Table E.241 Constraints and Portion of the Predictive Mean Vector for Health Insurance (UPMN), Any Other Health Insurance

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26 or Older
1	Missing ANYOTHER	497	None	1. HI1	1: 47	1: 216	1,2: 234

UPMN = univariate predictive mean neighborhood.

¹ The predictive mean vector components are defined by the following:

1. HI1 = P(Respondent has any other health insurance)

E.8 Roster Pair Variables

The tables in this section present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the pair variables. [Tables E.242](#) through [E.245](#) correspond to the first stage of imputation, where the pair relationship variable (IRPRREL) is created. [Tables E.246](#) through [E.262](#) are for the second stage, imputation of the multiplicity variables. Finally, [Tables E.263](#) through [E.279](#) correspond to the third stage, where the household counts are imputed.

There are a few instances where variable names are used without description for the purposes of brevity. In these cases, the variables are defined in a table at the beginning of the section in which they are used.

E.8.1 Stage One: Pair Relationship

Table E.242 Variables Used in Constraints for Stage One, Pair Relationship

Variable	Value	Constraint
MARIT1	1	At least one pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
MARIT2	2	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	1	One pair member has a marital status of "Married" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
MARIT3	5	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	4	One pair member has a marital status of "Married," and the other has a marital status of "Widowed" or "Divorced or Separated," neither of which was imputed or logically assigned
	3	One pair member has a marital status of "Married," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	2	Both pair members have a marital status of "Widowed" or "Divorced or Separated" that was not imputed or logically assigned
	1	One pair member has a marital status of "Widowed" or "Divorced or Separated," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	0	Both pair members either have a marital status of "Never Married" or were younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned

Table E.243 Logical Constraints for Stage One, Pair Relationship

Constraint #	Logical Constraint
LogC1	If recipient is a spouse-spouse pair, with or without children, then donor must be a spouse-spouse pair either with or without children
LogC2	If recipient could be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest
LogC3	If recipient could be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest
LogC4	If recipient could be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest
LogC5	If recipient could be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest
LogC6	If recipient could be either a sibling-sibling pair where one sibling is aged 12 to 17 and the other is aged 18 to 25 or some other clearly identifiable pair that is not of interest, then donor must be either a sibling-sibling (12-17/18-25) pair or some other clearly identifiable pair that is not of interest
LogC7	If recipient could be either a sibling-sibling pair where the siblings are not in either age range of interest or some other clearly identifiable pair that is not of interest, then donor must be a sibling-sibling pair where the siblings are not in an age range of interest or some other clearly identifiable pair that is not of interest
LogC8	If recipient could be either a spouse-spouse pair with children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with children or some other clearly identifiable pair that is not of interest
LogC9	If recipient could be either a spouse-spouse pair without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair without children or some other clearly identifiable pair that is not of interest
LogC10	If recipient could be a spouse-spouse pair with or without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with or without children or some other clearly identifiable pair that is not of interest
LogC11	If recipient could be either a grandparent-grandchild pair or another clearly identifiable pair that is not of interest, then donor must be a grandparent-grandchild pair or some other clearly identifiable pair that is not of interest

Table E.244 Likeness Constraints for Stage One, Pair Relationship

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is not a possible parent-child pair, and recipient has at least one parent in the household, then donor must also have at least one parent in the household
LikC3	If recipient is not a possible parent-child pair, and recipient has no parents in the household, then donor must also not have any parents in the household
LikC4	If the number of children aged 0 to 18 ¹ is nonmissing for both recipient and donor, then the two values must be equal
LikC5	If the recipient pair members are of the same gender, then the donor pair members must be of the same gender. If the recipient pair members are of a different gender, then the donor pair members must also be of a different gender.
LikC6	The age of the donor's younger pair member must be the same age as the recipient's younger pair member
LikC7	If donor is a spouse-spouse pair, with or without children, then donor's younger pair member must be the same age as recipient's younger pair member
LikC8	If both donor and recipient have pair members of different ages, then donor's younger pair member must be the same age as recipient's younger pair member
LikC9	The younger pair member of both donor and recipient must fall within the same age group: 21-25, 26-34, 35-49, or 50 or older
LikC10	The older pair member of both donor and recipient must fall within the same age group: 26-34, 35-49, or 50 or older
LikC11	If recipient pair members are in the same age group (21-25, 26-34, 35-49, or 50 or older), then donor cannot be a parent-child pair where the child is aged 21 or older
LikC12	Neither of the donor's pair members can have had an imputed marital status
LikC13	If donor is a sibling-sibling (12-14/15-17) pair, then neither of donor's pair members can have had an imputed marital status
LikC14	If neither of the recipient's pair members' marital status was imputed or if recipient has MARIT1 = 1, then donor must have the same value for MARIT1 as recipient
LikC15	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT2 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC16	If donor is a sibling-sibling (12-14/15-17) pair, then apply Likeness Constraint 12
LikC17	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT3 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC18	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value of MARIT3 as the recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT1 = 1, then donor must also have had MARIT1 = 1.

¹ For age group pairs 3 and 4, this constraint is on the number of children aged 0 to 11.

Table E.245 Constraints and Portion of the Predictive Mean Vector for Stage One, Pair Relationship

#	Age Group Pair ¹	Total Number of Cases	Logical Constraints	Predictive Mean Vector ²	Likeness Constraints: Number of Cases
1	One pair member is aged 12 to 14, the other is aged 15 to 17	13	None	1. SIB	1-3: 13
2	One pair member is aged 12 to 14, the other is aged 18 to 25	6	2,6	1. SIB	1-3,12,14: 6
3	Both pair members are aged 15 to 17	9	1,3,7-11	1. SIB	1-5,8,13,16: 3 2-5,8,12,15: 6
4	One pair member is aged 15 to 17, the other is aged 18 to 25	24	1,3,6-11	1. SIB	1-6,12,17: 9 2-6,12,17: 14 2-6,12,15: 0 2-6: 0 2-5: 1
5	One pair member is aged 18 to 20, the other is aged 18 to 25	28	1,4,7-10	1. SPOUSE1 2. SPOUSE2	1-5,12,17: 2 2-5,12,17: 26
6	Both pair members are aged 21 to 25	41	1,5,7-10	1. SPOUSE1 2. SPOUSE2	1-5,12,15: 8 2-5,12,15: 32 2-5: 1
7	One pair member is aged 12 to 14, the other is aged 26 or older	7	2,7,11	1. PC	1-3,10,12,14: 3 2,3,10,12,14: 4
8	One pair member is aged 15 to 17, the other is aged 26 or older	5	3,7-11	1. PC	1-5,7,10,12,18: 3 2-5,7,10,12,18: 1 2-5,7,12,18: 1
9	One pair member is aged 18 to 20, the other is aged 26 or older	14	1,4,7-11	1. PC	1-5,10,12,18: 3 2-5,10,12,18: 11
10	One pair member is aged 21 or older, the other is aged 26 or older	119	1,5,7-11	1. SPOUSE1 2. SPOUSE2	1-5,9-12,15: 21 2-5,9-12,15: 93 2-5,11,12,15: 5
11	Both pair members are aged 12 to 17	2	None	1. SIB	1-3: 2

¹ Because there was only one variable to be imputed, PAIRREL, there was only one missingness pattern. However, as the predictive mean vector and constraints vary by age group, the age groups are presented as missingness patterns in this table, and the column heading has been changed accordingly.

² The predictive mean vector components are defined by the following:

1. SIB = P(the pair relationship is sibling-sibling)
2. SPOUSE1 = P(the pair relationship is spouse-spouse, with children)
3. SPOUSE2 = P(the pair relationship is spouse-spouse, without children)
4. PC = P(the pair relationship is parent-child)

E.8.2 Stage Two: Parent-Child, Child Focus Multiplicity Counts

Table E.246 Likeness Constraints for Parent-Child, Child Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor's pair relationship must match recipient's pair relationship
LikC3	Donor and recipient must have the same household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC5	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC6	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's value must match
LikC7	If recipient's count of household members aged 26 or older is nonmissing, then donor's value must match
LikC8	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's value must match
LikC9	If recipient's count of household members aged 50 or older is nonmissing, then donor's value must match

Table E.247 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Child Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Completely missing	81	None	1. PRED1	1-6,8,9: 33 2-6,8,9: 40 2,3,6,8,9: 3 2,7: 5

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of parents of a child aged 12 to 20 who is a member of a parent-child pair.

E.8.3 Stage Two: Parent-Child, Parent Focus Multiplicity Counts

Table E.248 Logical Constraints for Parent-Child, Parent Focus Multiplicity Counts

Constraint #	Logical Constraint
LogC1	Donor's parent-child (12-14) parent focus multiplicity count must fall within recipient's bounds
LogC2	Donor's parent-child (12-17) parent focus multiplicity count must fall within recipient's bounds
LogC3	Donor's parent-child (12-20) parent focus multiplicity count must fall within recipient's bounds
LogC4	Donor's parent-child (15-17) parent focus multiplicity count must fall within recipient's bounds

Table E.249 Likeness Constraints for Parent-Child, Parent Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor's pair relationship must match recipient's pair relationship
LikC3	Donor and recipient must have the same household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC5	If recipient is a parent-child (12-14) pair, and recipient's count of household members aged 12 to 14 is nonmissing, then donor's value must match
LikC6	If recipient is a parent-child (12-14) or (15-17) pair, and recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC7	If recipient's count of household members aged 12 to 20 is nonmissing, then donor's value must match
LikC8	If recipient is a parent-child (15-17) pair, and recipient's count of household members aged 15 to 17 is nonmissing, then donor's value must match
LikC9	If recipient is a parent-child (12-14) pair, then donor and recipient must have the same value for the parent-child (12-14) parent focus multiplicity count
LikC10	If recipient is a parent-child (12-14) or (12-17) pair, then donor and recipient must have the same value for the parent-child (12-17) parent focus multiplicity count
LikC11	Donor and recipient must have the same value for the parent-child (12-20) parent focus multiplicity count
LikC12	If recipient is a parent-child (15-17) pair, then donor and recipient must have the same value for the parent-child (15-17) parent focus multiplicity count

Table E.250 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Parent Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing 12-14, 12-17, and 12-20 multiplicity counts	3	1-3	1. PRED1	1-8: 3
2	Missing 12-17, 12-20, and 15-17 multiplicity counts	8	2-4	1. PRED1	1-8: 5 2-8: 0 2,3,5-8: 0 2,5-8: 3
3	Missing 12-14 and 12-17 multiplicity counts	0	1,2	1. PRED1	No Cases
4	Missing 12-14 and 12-20 multiplicity counts	0	1,3	1. PRED1	No Cases
5	Missing 12-17 and 12-20 multiplicity counts	8	2,3	1. PRED1	1-9,12: 7 2-9,12: 0 2,3,5-9,12: 0 2,5-9,12: 1
6	Missing 12-17 and 15-17 multiplicity counts	0	2,4	1. PRED1	No Cases
7	Missing 12-20 and 15-17 multiplicity counts	0	3,4	1. PRED1	No Cases
8	Missing the 12-14 multiplicity count	1	1	1. PRED1	1-8,10,11: 1
9	Missing the 12-17 multiplicity count	0	2	1. PRED1	No Cases
10	Missing the 12-20 multiplicity count	16	3	1. PRED1	1-10,12: 12 2-10,12: 1 2,3,5-10,12: 1 2,5-10,12: 1 5-10,12: 1
11	Missing the 15-17 multiplicity count	0	4	1. PRED1	No Cases

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of children aged 12 to 20 belonging to a parent who is a member of a parent-child pair.

E.8.4 Stage Two: Sibling-Sibling (12-14/15-17) Older Focus Multiplicity Counts

Table E.251 Logical Constraints for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts

Constraint #	Logical Constraint
LogC1	Donor's sibling-sibling (12-14/15-17) older focus multiplicity count must fall within recipient's bounds

Table E.252 Likeness Constraints for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor and recipient must have the same household size
LikC3	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC4	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's value must match
LikC5	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match

Table E.253 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	3	1	1. PRED1	1-5: 2 2-5: 0 2,4: 0 4: 1

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of siblings aged 12 to 14 for a respondent aged 15 to 17 who is a member of a sibling-sibling pair.

E.8.5 Stage Two: Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

Table E.254 Logical Constraints for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

Constraint #	Logical Constraint
LogC1	Donor's sibling-sibling (12-14/15-17) younger focus multiplicity count must fall within recipient's bounds

Table E.255 Likeness Constraints for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor and recipient must have the same household size
LikC3	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC4	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's value must match
LikC5	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC6	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's value must match

Table E.256 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	3	1	1. PRED1	1-5: 3

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of siblings aged 15 to 17 for a respondent aged 12 to 14 who is a member of a sibling-sibling pair.

E.8.6 Stage Two: Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

Table E.257 Logical Constraints for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

Constraint #	Logical Constraint
LogC1	Donor's sibling-sibling (12-17/18-25) older focus multiplicity count must fall within recipient's bounds

Table E.258 Likeness Constraints for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor and recipient must have the same household size
LikC3	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC4	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC5	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's value must match

Table E.259 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	9	1	1. PRED1	1-5: 5 2-5: 3 2,4: 0 4: 1

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of siblings aged 12 to 17 for a respondent aged 18 to 25 who is a member of a sibling-sibling pair.

E.8.7 Stage Two: Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

Table E.260 Logical Constraints for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

Constraint #	Logical Constraint
LogC1	Donor's sibling-sibling (12-17/18-25) younger focus multiplicity count must fall within recipient's bounds

Table E.261 Likeness Constraints for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor and recipient must have the same household size
LikC3	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC4	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC5	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's value must match

Table E.262 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	11	1	1. PRED1	1-5: 8 2-5: 0 2,5: 0 5: 3

¹ The predictive mean vector components are defined by the following:

1. PRED1 is the predicted mean for the number of siblings aged 18 to 25 for a respondent aged 12 to 17 who is a member of a sibling-sibling pair.

E.8.8 Stage Three: Parent-Child Household Counts**Table E.263 Logical Constraints for Parent-Child Household Counts**

Constraint #	Logical Constraint
LogC1	Donor's parent-child (12-14) parent count must be greater than 0
LogC2	Donor's parent-child (12-17) parent count must be greater than 0
LogC3	Donor's parent-child (12-20) parent count must be greater than 0
LogC4	Donor's parent-child (12-14) child count must be greater than 0
LogC5	Donor's parent-child (12-17) child count must be greater than 0
LogC6	Donor's parent-child (12-20) child count must be greater than 0
LogC7	Donor's parent-child (12-14) child count must fall within recipient's bounds
LogC8	Donor's parent-child (12-17) child count must fall within recipient's bounds
LogC9	Donor's parent-child (12-20) child count must fall within recipient's bounds
LogC10	Donor's parent-child (12-14) parent count must fall within recipient's bounds
LogC11	Donor's parent-child (12-17) parent count must fall within recipient's bounds
LogC12	Donor's parent-child (12-20) parent count must fall within recipient's bounds

Table E.264 Likeness Constraints for Stage Three, Household Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient lives in a multi-family home, then donor must also live in a multi-family home
LikC3	If recipient is part of a pair, then donor must also be part of a pair. If recipient is not part of a pair, then donor must also not be part of a pair.
LikC4	Donor's household size must be the same as recipient's household size
LikC5	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's count must match
LikC7	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's count must match
LikC8	If recipient's count of household members aged 12 to 20 is nonmissing, then donor's count must match
LikC9	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC10	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's count must match
LikC11	If recipient's count of household members aged 26 to 49 is nonmissing, then donor's count must match
LikC12	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's count must match
LikC13	If recipient's count of household members aged 50 or older is nonmissing, then donor's count must match
LikC14	If recipient could possibly have a parent-child (12-14) parent count of 0, then donor's and recipient's counts of household members aged 12 to 14 must match
LikC15	If recipient could possibly have a parent-child (12-17) parent count of 0, then donor's and recipient's counts of household members aged 12 to 17 must match
LikC16	If recipient could possibly have a parent-child (12-20) parent count of 0, then donor's and recipient's counts of household members aged 12 to 20 must match
LikC17	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC18	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC19	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC20	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC21	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC22	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing PPCP1220	22	3,12	1. PC1 2. PC2 3. H3	1-5,9,10,12,13,16: 0 2-5,9,10,12,13,16: 8 2,3,5,9,10,12,13,16: 0 2,3,11,16: 12 2,3,16: 2
2	Missing PPCP1217	3	2,11	1. PC1 2. PC2 3. H3	1-5,9,10,12,13,15: 0 2-5,9,10,12,13,15: 3
3	Missing PPCP1217 and PPCP1220	22	2,3,11,12	1. PC1 2. PC2 3. H3	1-5,9,10,12,13,15,16: 3 2-5,9,10,12,13,15,16: 13 2,3,5,9,10,12,13,15,16: 0 2,3,11,15,16: 4 2,3,15,16: 2
4	Missing PPCP1214	4	1,10	1. PC1 2. PC2 3. H3	1-5,9,10,12-14: 0 2-5,9,10,12-14: 4
5	Missing PPCP1214 and PPCP1220	0	1,3,10,12	1. PC1 2. PC2 3. H3	No Cases
6	Missing PPCP1214 and PPCP1217	2	1,2,10,11	1. PC1 2. PC2 3. H3	1-5,9,10,12-15: 0 2-5,9,10,12-15: 2
7	Missing PPCP1214, PPCP1217, and PPCP1220	39	1-3,10-12	1. PC1 2. PC2 3. H3	1-5,9,10,12-16: 4 2-5,9,10,12-16: 29 2,3,5,9,10,12-16: 2 2,3,11,14-16: 3 2,3,14-16: 1
8	Missing PPCC1220	19	6,9	1. PC1 2. PC2 3. H3	1-5,8,9,19: 2 2-5,8,9,19: 13 2,3,5,8,9,19: 1 2,3,8,22: 3

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
9	Missing PPCC1220 and PPCP1220	147	9,12	1. PC1 2. PC2 3. H3	1-5,8-10,12,13,16,19: 17 2-5,8-10,12,13,16,19: 4 2,3,5,8-10,12,13,16,19: 1 2,3,8,11,16,22: 93 2,3,8,16: 26 2,3,16: 4 2,3: 2
10	Missing PPCC1220 and PPCP1217	0	2,6,9,11	1. PC1 2. PC2 3. H3	No Cases
11	Missing PPCC1220, PPCP1217, and PPCP1220	0	2,9,11,12	1. PC1 2. PC2 3. H3	No Cases
12	Missing PPCC1220 and PPCP1214	0	1,6,9,10	1. PC1 2. PC2 3. H3	No Cases
13	Missing PPCC1220, PPCP1214, and PPCP1220	0	1,9,10,12	1. PC1 2. PC2 3. H3	No Cases
14	Missing PPCC1220, PPCP1214, and PPCP1217	0	1,2,6,9-11	1. PC1 2. PC2 3. H3	No Cases
15	Missing PPCC1220, PPCP1214, PPCP1217, and PPCP1220	0	1,2,9-12	1. PC1 2. PC2 3. H3	No Cases
16	Missing PPCC1217	0	5,8	1. PC1 2. PC2 3. H3	No Cases
17	Missing PPCC1217 and PPCP1220	0	3,5,8,12	1. PC1 2. PC2 3. H3	No Cases
18	Missing PPCC1217 and PPCP1217	0	8,11	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
19	Missing PPCC1217, PPCP1217, and PPCP1220	0	3,8,11,12	1. PC1 2. PC2 3. H3	No Cases
20	Missing PPCC1217 and PPCP1214	0	1,5,8,10	1. PC1 2. PC2 3. H3	No Cases
21	Missing PPCC1217, PPCP1214, and PPCP1220	0	1,3,5,8,10,12	1. PC1 2. PC2 3. H3	No Cases
22	Missing PPCC1217, PPCP1214, and PPCP1217	0	1,8,10,11	1. PC1 2. PC2 3. H3	No Cases
23	Missing PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	1,3,8,10-12	1. PC1 2. PC2 3. H3	No Cases
24	Missing PPCC1217 and PPCC1220	15	5,6,8,9	1. PC1 2. PC2 3. H3	1-5,7-9,18,19: 1 2-5,7-9,18,19: 8 2,3,5,7-9,18,19: 2 2,3,7,8,21,22: 0 2,3,7,8: 0 2,3: 3 1,2,4,5,7-9,18,19: 0 2,4,5,7-9,18,19: 0 2,5,7-9,18,19: 0 2,7,8,21,22: 0 2,7,8: 0 2: 1
25	Missing PPCC1217, PPCC1220, and PPCP1220	0	5,8,9,12	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
26	Missing PPCC1217, PPCC1220, and PPCP1217	1	6,8,9,11	1. PC1 2. PC2 3. H3	1-5,7-10,12,13,15,18,19: 0 2-5,7-10,12,13,15,18,19: 0 2,3,5,7-10,12,13,15,18,19: 0 2,3,7,8,11,15,21,22: 0 2,3,7,8,15: 0 2,3,15: 0 2,3: 1
27	Missing PPCC1217, PPCC1220, PPCP1217, and PPCP1220	125	8,9,11,12	1. PC1 2. PC2 3. H3	1-5,7-10,12,13,15,16,18,19: 6 2-5,7-10,12,13,15,16,18,19: 3 2,3,5,7-10,12,13,15,16,18,19: 0 2,3,7,8,11,15,16,21,22: 55 2,3,7,8,15,16: 27 2,3,15,16: 1 2,3: 20 1,2,4,5,7-10,12,13,15,16,18,19: 1 2,4,5,7-10,12,13,15,16,18,19: 0 2,5,7-10,12,13,15,16,18,19: 0 2,7,8,11,15,16,21,22: 6 2,7,8,15,16: 3 2,15,16: 3
28	Missing PPCC1217, PPCC1220, and PPCP1214	1	1,5,6,8-10	1. PC1 2. PC2 3. H3	1-5,7-10,12-14,18,19: 0 2-5,7-10,12-14,18,19: 0 2,3,5,7-10,12-14,18,19: 0 2,3,7,8,11,14,21,22: 0 2,3,7,8,14: 0 2,3,14: 1
29	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	1,5,8-10,12	1. PC1 2. PC2 3. H3	No Cases
30	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1217	0	1,6,8-11	1. PC1 2. PC2 3. H3	No Cases
31	Missing PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	0	1,8-12	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
32	Missing PPCC1214	0	4,7	1. PC1 2. PC2 3. H3	No Cases
33	Missing PPCC1214 and PPCP1220	0	3,4,7,12	1. PC1 2. PC2 3. H3	No Cases
34	Missing PPCC1214 and PPCP1217	0	2,4,7,11	1. PC1 2. PC2 3. H3	No Cases
35	Missing PPCC1214, PPCP1217, and PPCP1220	0	2-4,7,11,12	1. PC1 2. PC2 3. H3	No Cases
36	Missing PPCC1214 and PPCP1214	1	7,10	1. PC1 2. PC2 3. H3	1-6,9,10,12-14,17: 0 2-6,9,10,12-14,17: 0 2,3,5,6,9,10,12-14,17: 0 2,3,6,11,14,20: 0 2,3,6,14: 1
37	Missing PPCC1214, PPCP1214, and PPCP1220	0	3,7,10,12	1. PC1 2. PC2 3. H3	No Cases
38	Missing PPCP1214 and PPCP1217	0	2,7,10,11	1. PC1 2. PC2 3. H3	No Cases
39	Missing PPCP1214, PPCP1217, and PPCP1220	0	2,3,7,10-12	1. PC1 2. PC2 3. H3	No Cases
40	Missing PPCC1214 and PPCC1220	0	4,6,7,9	1. PC1 2. PC2 3. H3	No Cases
41	Missing PPCC1214, PPCC1220, and PPCP1220	0	4,7,9,12	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
42	Missing PPCC1214, PPCC1220, and PPCP1217	0	2,4,6,7,9,11	1. PC1 2. PC2 3. H3	No Cases
43	Missing PPCC1214, PPCC1220, PPCP1217, and PPCP1220	0	2,4,7,9,11,12	1. PC1 2. PC2 3. H3	No Cases
44	Missing PPCC1214, PPCC1220, and PPCP1214	0	6,7,9,10	1. PC1 2. PC2 3. H3	No Cases
45	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1220	0	7,9,10,12	1. PC1 2. PC2 3. H3	No Cases
46	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1217	0	2,6,7,9-11	1. PC1 2. PC2 3. H3	No Cases
47	Missing PPCC1214, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	0	2,7,9-12	1. PC1 2. PC2 3. H3	No Cases
48	Missing PPCC1214 and PPCC1217	0	4,5,7,8	1. PC1 2. PC2 3. H3	No Cases
49	Missing PPCC1214, PPCC1217, and PPCP1220	0	3-5,7,8,12	1. PC1 2. PC2 3. H3	No Cases
50	Missing PPCC1214, PPCC1217, and PPCP1217	0	4,7,8,11	1. PC1 2. PC2 3. H3	No Cases
51	Missing PPCC1214, PPCC1217, PPCP1217, and PPCP1220	0	3,4,7,8,11,12	1. PC1 2. PC2 3. H3	No Cases
52	Missing PPCC1214, PPCC1217, and PPCP1214	0	5,7,8,10	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
53	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1220	0	3,5,7,8,10,12	1. PC1 2. PC2 3. H3	No Cases
54	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1217	1	7,8,10,11	1. PC1 2. PC2 3. H3	1-7,9,10,12-15,17,18: 0 2-7,9,10,12-15,17,18: 0 2,3,5-7,9,10,12-15,17,18: 0 2,3,6,7,11,14,15,20,21: 1
55	Missing PPCC1214, PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	3,7,8,10-12	1. PC1 2. PC2 3. H3	No Cases
56	Missing PPCC1214, PPCC1217, and PPCC1220	16	4-9	1. PC1 2. PC2 3. H3	1-9,17-19: 1 2-9,17-19: 12 2,3,5-9,17-19: 1 2,3,6-8,20-22: 2
57	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1220	0	4,5,7-9,12	1. PC1 2. PC2 3. H3	No Cases
58	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1217	0	4,6-9,11	1. PC1 2. PC2 3. H3	No Cases
59	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1217, and PPCP1220	0	4,7-9,11,12	1. PC1 2. PC2 3. H3	No Cases
60	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1214	0	5-10	1. PC1 2. PC2 3. H3	No Cases
61	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	5,7-10,12	1. PC1 2. PC2 3. H3	No Cases

Table E.265 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
62	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1217	1	6-11	1. PC1 2. PC2 3. H3	1-10,12-15,17-19: 0 2-10,12-15,17-19: 0 2,3,5-10,12-15,17-19: 0 2,3,6-8,11,14,15,20-22: 0 2,3,6-8,14,15: 0 2,3,14,15: 0 2,3: 1
63	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	141	7-12	1. PC1 2. PC2 3. H3	1-10,12-19: 3 2-10,12-19: 1 2,3,5-10,12-19: 0 2,3,6-8,11,14-16,20-22: 32 2,3,6-8,14-16: 54 2,3,14-16: 2 2,3: 38 1,2,4-10,12-19: 0 2,4-10,12-19: 0 2,5-10,12-19: 1 2,6-8,11,14-16,20-22: 3 2,6-8,14-16: 2 2,14-16: 5

¹ The predictive mean vector components are defined by the following:

1. H3 = Predicted mean count of children aged 12 to 20 with parents in the household
2. PC1 = P(there is one parent with at least one child aged 12 to 20 in the household)
3. PC2 = P(there are two parents with at least one child aged 12 to 20 in the household)

E.8.9 Stage Three: Household Sibling Count (12-14/15-17), Older Focus**Table E.266 Logical Constraints for Household Sibling Count (12-14/15-17), Older Focus**

Constraint #	Logical Constraint
LogC1	Donor's count of older siblings aged 15 to 17 must fall within recipient's bounds
LogC2	Donor's count of older siblings aged 15 to 17 cannot be larger than recipient's nonmissing count of household members aged 15 to 17

Table E.267 Likeness Constraints for Household Sibling Count (12-14/15-17), Older Focus

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is part of a pair, then donor must also be part of a pair. If recipient is not part of a pair, then donor must also not be part of a pair.
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's count must match
LikC5	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's count must match
LikC7	Donor's and recipient's counts of household members aged 12 to 14 must either both be 0 or both be positive
LikC8	Donor's and recipient's counts of household members aged 15 to 17 must either both be 0 or both be positive
LikC9	If recipient's count of household members aged 12 to 14 is missing, then donor and recipient must have the same screener count of household members aged 12 to 14
LikC10	If recipient's count of household members aged 15 to 17 is missing, then donor and recipient must have the same screener count of household members aged 15 to 17
LikC11	Donor and recipient must have parent-child (12-14) parent counts that are both 0 or both positive
LikC12	Donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive
LikC13	If recipient's counts of household members aged 12 to 14 and 15 to 17 are both missing, then donor and recipient must have parent-child (12-14 and 12-17) parent counts that are both 0 or both positive

Table E.268 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-14/15-17), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	55	1,2	1. H1	1-6,9-12: 16 2-6,9-12: 22 2,5,6,9-12: 12 2,7,8,13: 5

¹ The predictive mean vector components are defined by the following:

1. H1 is the predicted mean count of older siblings aged 15 to 17.

E.8.10 Stage Three: Household Sibling Count (12-17/18-25), Older Focus**Table E.269 Logical Constraints for Household Sibling Count (12-17/18-25), Older Focus**

Constraint #	Logical Constraint
LogC3	Donor's count of older siblings aged 18 to 25 must fall within recipient's bounds
LogC4	Donor's count of older siblings aged 18 to 25 cannot be larger than recipient's nonmissing count of household members aged 18 to 25

Table E.270 Likeness Constraints for Household Sibling Count (12-17/18-25), Older Focus

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is part of a pair, then donor must also be part of a pair. If recipient is not part of a pair, then donor must also not be part of a pair.
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's count must match
LikC5	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC7	Donor's and recipient's counts of household members aged 12 to 17 must either both be 0 or both be positive
LikC8	Donor's and recipient's counts of household members aged 18 to 25 must either both be 0 or both be positive
LikC9	If recipient's count of household members aged 12 to 17 is missing, then donor and recipient must have the same screener count of household members aged 12 to 17
LikC10	If recipient's count of household members aged 18 to 25 is missing, then donor and recipient must have the same screener count of household members aged 18 to 25
LikC11	Donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive
LikC12	If recipient's counts of household members aged 12 to 17 and 18 to 25 are both missing, then donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive

Table E.271 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-17/18-25), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	148	1,2	1. H2	1-6,9-11: 33 2-6,9-11: 69 2,5,6,9-11: 31 2,7,8,12: 15

¹ The predictive mean vector components are defined by the following:
1. H2 is the predicted mean count of older siblings aged 18 to 25.

E.8.11 Stage Three: Household Spouse-Spouse Count

Table E.272 Variables Used in Constraints for the Household Spouse-Spouse Count

Variable	Value	Constraint
Males XXXX ¹	0	No household members in the specified age group are male
	1	All household members in the specified age group are male
	2	Some household members in the specified age group are male
	3	There are no household members in the specified age group

¹ XXXX denotes the 4-digit age range (e.g., 1217 denotes the age range 12 to 17).

Table E.273 Logical Constraints for the Household Spouse-Spouse Count

Constraint #	Logical Constraint
LogC1	Donor's count of spouse-spouse pairs must fall within recipient's bounds

Table E.274 Likeness Constraints for Household Spouse-Spouse Count

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient lives in a multi-family home, then donor must also live in a multi-family home
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient is part of a pair, then donor must also be part of a pair. If recipient is not part of a pair, then donor must also not be part of a pair.
LikC5	If recipient's count of household members aged 15 or older is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC7	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's count must match
LikC8	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's count must match
LikC9	If recipient's count of household members aged 50 or older is nonmissing, then donor's count must match
LikC10	If recipient's value of MALES15P is nonmissing, then donor's value of MALES15P must match
LikC11	If recipient's count of household members aged 15 or older is positive, then donor and recipient must have the same count of household members aged 15 or older and the same values for MALES15P
LikC12	If recipient's value of MALES1825 is nonmissing, then donor's value of MALES1825 must match
LikC13	If recipient's value of MALES1834 is nonmissing, then donor's value of MALES1834 must match
LikC14	If recipient's value of MALES2634 is nonmissing, then donor's value of MALES2634 must match
LikC15	If recipient's value of MALES2649 is nonmissing, then donor's value of MALES2649 must match
LikC16	If recipient's value of MALES2649 is nonmissing and not equal to 3, then donor's value of MALES2649 must match
LikC17	If recipient's value of MALES3549 is nonmissing, then donor's value of MALES3549 must match
LikC18	If recipient's value of MALES50P is nonmissing, then donor's value of MALES50P must match

Table E.275 Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	1,457	1	1. SPOUSE1 2. SPOUSE2	1-11,13,15,18,19: 454 2-11, 13,15,18,19: 473 2,4,10,12,13,15,18,19: 133 2,4,12,14,16,19: 293 2,4,12,17: 90 2,4,12: 5 2,4: 9

- ¹ The predictive mean vector components are defined by the following:
1. SPOUSE1 = P(the number of spouse-spouse pairs in the household is 0)
 2. SPOUSE2 = P(the number of spouse-spouse pairs in the household is 1)

E.8.12 Stage Three: Household Spouse-Spouse with Children Count**Table E.276 Variables Used in Constraints for Household Spouse-Spouse with Children Count**

Variable	Value	Constraint
Males XXXX ¹	0	No household members in the specified age group are male
	1	All household members in the specified age group are male
	2	Some household members in the specified age group are male
	3	There are no household members in the specified age group

- ¹ XXXX denotes the 4-digit age range (e.g., 1217 denotes the age range 12 to 17).

Table E.277 Logical Constraints for Household Spouse-Spouse with Children Count

Constraint #	Logical Constraint
LogC1	Donor's count of spouse-spouse pairs with children must fall within recipient's bounds

Table E.278 Likeness Constraints for Household Spouse-Spouse with Children Count

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient lives in a multi-family home, then donor must also live in a multi-family home
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient is part of a pair, then donor must also be part of a pair. If recipient is not part of a pair, then donor must also not be part of a pair.
LikC5	If recipient's count of household members aged 15 or older is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC7	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's count must match
LikC8	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's count must match
LikC9	If recipient's count of household members aged 50 or older is nonmissing, then donor's count must match
LikC10	Donor's and recipient's counts of household members aged 0 to 17 must both be 0, 1, or greater than 1
LikC11	Donor's and recipient's counts of household members aged 0 to 17 must either both be 0 or both be positive
LikC12	If recipient's value of MALES15P is nonmissing, then donor's value of MALES15P must match
LikC13	If recipient's count of household members aged 15 or older is positive, then donor and recipient must have the same count of household members aged 15 or older and the same values for MALES15P
LikC14	If recipient's value of MALES1825 is nonmissing, then donor's value of MALES1825 must match
LikC15	If recipient's value of MALES1834 is nonmissing, then donor's value of MALES1834 must match
LikC16	If recipient's value of MALES2634 is nonmissing, then donor's value of MALES2634 must match
LikC17	If recipient's value of MALES2649 is nonmissing, then donor's value of MALES2649 must match
LikC18	If recipient's value of MALES2649 is nonmissing and not equal to 3, then donor's value of MALES2649 must match
LikC19	If recipient's value of MALES3549 is nonmissing, then donor's value of MALES3549 must match
LikC20	If recipient's value of MALES50P is nonmissing, then donor's value of MALES50P must match

Table E.279 Constraints and Portion of the Predictive Mean Vector for the Household Spouse-Spouse with Children Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector ¹	Likeness Constraints: Number of Cases
1	Missing	603	1	1. SPOUSE3	1-10,12,14,16,19,20: 172 2-10,12,14,16,19,20: 184 2,4,10,13,14,16,19,20: 72 2,4,11,13,15,17,20: 117 2,4,11,13,18: 47 2,4,11,13: 4 1-3,5-10,12,14,16,19,20: 0 2,3,5-10,12,14,16,19,20: 0 2,10,13,14,16,19,20: 0 2,11,13,15,17,20: 1 2,11,13,18: 0 2,11,13: 0 2,11,13*: 6

*A donor could not be found for these cases. Imputed values were randomly assigned within the bounds created during the editing process.

¹ The predictive mean vector components are defined by the following:

1. SPOUSE3 = P(there is at least one spouse-spouse pair with children in the household)

Appendix F: Quality Control Measures Used in the Imputation Procedures

F.1 Introduction

For the 2023 National Survey on Drug Use and Health (NSDUH), several quality control (QC) measures were implemented for the imputation of demographic, substance use, substance use disorder, substance use treatment, adult mental health, mental health services utilization, income, health insurance, nicotine dependence, household composition (roster), and pair variables. These QC measures spanned the following three basic steps within the predictive mean neighborhood (PMN) and modified predictive mean neighborhood (modPMN) imputation methodologies: (1) response propensity adjustment, (2) prediction modeling, and (3) imputation.

Checklists for the imputation of these variables were used during imputation processing and serve as formal documentation of the QC measures that were implemented. Most of these checks required both a technician check, where the computer program technicians entered their name and the date the check was performed, and an independent reviewer check, where independent reviewers performed a check of the same items. The reviewer check ensured greater quality in the imputation procedures. Additionally, some checklist entries required the technician to document the procedures that were taken to run the programs, such as listing the variables that were dropped from the model in order to achieve model convergence.

The checks involved in each of the steps are described in detail in the following sections. Variable-specific consistency checks were often created in addition to the checks of the PMN and modPMN steps. However, they are not discussed in this appendix. The QC measures that were used in the editing process or the delivery of variables to other NSDUH teams also are not discussed.

F.2 Step 1: Response Propensity Adjustment

The first step of the PMN and modPMN methodologies is a response propensity adjustment. The item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special circumstance of the generalized exponential model (GEM)¹ developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

For PMN, the sampling weights are adjusted for item nonresponse so that the item respondent weights are representative of the entire domain of interest. These adjusted sampling weights are used in the subsequent prediction modeling and hot-deck imputation steps. For modPMN, the adjusted sampling weights are not used in subsequent steps; however, the respondents and

¹ The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of the *2023 National Survey on Drug Use and Health Methodological Resource Book, Section 11: Person-Level Sampling Weight Calibration* report (Center for Behavioral Health Statistics and Quality, 2025b).

nonrespondents are segregated into five bins of roughly equal size based on their predicted response propensities from the model. The prediction modeling and imputation are then performed separately for each bin.

The following QC measures were conducted in Step 1:

- The output of the response propensity modeling program was checked for singularities. Any singularities that occurred were investigated, and the model was corrected by removing the correlated covariates from the model. For response propensity modeling programs that used the automated model-fitting techniques (see Chapter 3), singularities were removed from the model automatically by removing the covariates causing the singularity. The QC check then involved comparing the list of singularities to the list of covariates in the output to ensure the variables had been removed.
- Similarly, the automated model-fitting techniques identified and resolved data separation issues by either collapsing the covariate or removing the covariate. A QC check ensured the variables flagged as causing an unresolvable separation issue did not appear in the model output. Additional QC checks ensured that no variables remaining in the model would cause separation issues and that the automated model-fitting techniques did not remove all categorical variables from the model.
- The values of the convergence flags were checked to see whether the GEM converged.
- The current distance to root was checked to make sure it was close to 0, indicating that GEM converged.
- When variables were reduced from the original model, the remaining levels of variables were checked to ensure appropriateness. An example of this check was to determine whether the base variables or lower order terms were present when interactions or higher order terms existed (e.g., “age” and “age squared” must be in the model if “age cubed” is in the model).
- An indicator was calculated in the response propensity program that measured the maximum adjustment to the weights. In most situations, the adjusted weights resembled the original design weights. If the maximum adjustment was too high (usually greater than 3.00), this was likely due to an overspecified model, where the adjustment was not performing at an optimum level. Large maximum adjustments were investigated and corrected, if possible, by removing extraneous variables from the models so that any final weight adjustment applied to a respondent was within acceptable bounds.
- After the weights were adjusted, the ratio of the maximum adjusted weight to the mean adjusted weight (called the “mmratio”) was computed for both the unadjusted and adjusted weights to monitor the change in variation among the weights. Any relative change in the mmratio value that was greater than 25 percent was noted in the response propensity program checklist.

- The unequal weighting effect (UWE) was checked before and after adjustment to ensure that there was no significant variance increase due to the nonresponse adjustment. The difference in the UWE after-adjustment value should be no more than 20 percent of the UWE before-adjustment value. The difference was fairly small in most instances, and any difference greater than 20 percent was investigated and corrected, if possible.
- The weighted totals for the independent variables in the model were compared both before and after the adjustment. If these weighted totals were equal, the adjustment procedures worked properly.
- For modPMN, the five bins were checked to ensure that approximately 20 percent of the domain members was assigned to each bin and that the larger the bin number, the larger the number of nonrespondents (i.e., a larger bin number indicates a lower estimated response propensity from the model). When the number of respondents within at least one bin was lower than a specified threshold (i.e., fewer than 10) and the bin(s) below this threshold also contained at least one nonrespondent, the five response propensity bins were collapsed. In this situation, QC checks ensured that the conditions for collapsing bins had been met and that the collapsing procedures correctly placed all domain members into a single bin.

F.3 Step 2: Prediction Modeling

For PMN, the prediction modeling step, using the adjusted weights from Step 1, calculates predicted means, which are then used in the hot-deck step(s) to create neighborhoods and select donors. For modPMN, the prediction modeling step is conducted separately within each of the five response propensity bins from Step 1, or within a single bin if the bins are collapsed, using the unadjusted sampling weights. The predicted means from this step are then used in the imputation step. The dependent variable in the model is usually the variable that is undergoing imputation, or some transformation of that variable. Each model is built using only those respondents within the domain with complete responses for that item. Predicted means are then calculated for all the domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, binomial or multinomial logistic regression models are used for the prediction models depending on the number of levels for the dependent variable. For continuous variables, linear regression models are fit. For count variables, Poisson regression models are used. For response variables that are proportions, a logit transformation is applied to the proportion, and a linear regression model is utilized. For continuous nonnegative random variables (e.g., finer categories of income), a failure time model is used. The models are discussed in detail in the imputation sections of Chapters 4 through 8 and Chapters 10 through 14 of this report.

The following QC measures were conducted in Step 2:

- All models were checked for singularities and collinearities. Any singularities that occurred were investigated, and the model was corrected. For models that used the automated model-fitting techniques, singularities were removed from the model automatically by removing the covariates causing the singularity. The QC check then involved comparing the list of singularities to the list of covariates in the output to ensure the variables had been removed.
- Similarly, the automated model-fitting techniques identified and resolved data separation issues by either collapsing the covariate or removing the covariate. A QC check ensured the variables flagged as causing an unresolvable separation issue did not appear in the model output. Additional QC checks ensured that no variables remaining in the model would cause separation issues and that the automated model-fitting techniques did not remove all categorical variables from the model.
- For Poisson regression models, failure time models, and logistic models, the output was checked to ensure convergence had occurred.
- The log file was checked for “data warning” messages or other SUDAAN®-specific errors.² Under PMN, if there was a “data warning” message in the log, independent variables were dropped until the modeling procedure in SUDAAN produced no warning messages. Under modPMN, many of these warning messages are ignored under the following conditions: the algorithm converges, reasonable estimates of regression coefficients are produced, and reasonable test statistics associated with the regression coefficients are produced.
- When variables were manually reduced from the original model, the remaining levels of variables were checked to ensure appropriateness. Similar to the response propensity model, if the main variable was dropped, its interaction variables were also dropped.
- For models that used the automated model-fitting techniques, sets of covariates were automatically dropped when the overall Wald F test statistic for a related set of covariates was less than 1. When this occurred, checks ensured that (1) interaction terms were removed before main effects, and (2) only those variables with a Wald F test statistic less than 1, or a higher order interaction term if the main effect was flagged, were removed from the model.
- Output was checked to verify that everything worked properly in the regression model.
- A check on the predicted means from the model was created to ensure that each respondent in the domain had a valid nonmissing predicted mean.

F.4 Step 3: Imputation

The predicted means from Step 2 were used to determine the final assignments of imputed values in Step 3. Final imputed values were assigned using either a hot-deck step to select a donor or a collocated stochastic imputation process using the nonrespondent’s predicted mean(s) from Step 2 (see Chapter 3 for more details).

² Details can be found in the *SUDAAN® Language Manual, Release 11.0.1* (RTI International, 2013).

The QC measures in this step had two objectives: (1) ensure that the imputed values were consistent with preexisting nonmissing values, and (2) ensure that the imputed values were assigned as intended. Many variables had logical constraints or restrictions placed on the final imputed value that had to be adhered to in order to prevent logical inconsistencies. Specific QC checks were implemented to ensure this; however, only the more general checks that apply to all imputations are listed in this appendix. For more detail on logical constraints and restrictions placed on imputed values, refer to Appendix E.

The following QC measures were conducted in Step 3:

- Unusual imputed values were noted. If the imputed value was equivalent to one of the standard NSDUH missing value codes, this signaled a failure in the imputation process (e.g., a suitable donor could not be found), and measures were required (e.g., likeness constraints were loosened) to revise the programs so that a final imputed value could be assigned. If the imputed value was otherwise unusual, the imputation process was examined to ensure that no errors occurred.
- For variables imputed using a hot-deck step, the number imputed within various levels of restrictiveness of the likeness constraints was noted.³
- For variables imputed using a hot-deck step, the frequency of the variable WORKED was checked to ensure that no values were equal to 0. Values greater than 0 signified that the imputation procedure was able to find a donor.
- The distribution of edited variables was compared with the distribution of imputed variables to make sure that each imputed value was within the appropriate range corresponding to the value of the edited variable.
- The imputed values were crossed with the imputation indicators to ensure that the indicators were created correctly.
- After imputation was implemented, the distribution of values for nonrespondents was checked against the distribution of values for all respondents to ensure the similarity of these two subgroups.
- It was necessary to ensure that everyone to whom the variable did not apply received a skip code for the final imputed variable. For example, all those in the 12-14 age group should have a "Not Applicable (N/A)" value of 99 for the imputation-revised marital status variable IRMARIT.
- Any missing values were noted. This occurred when the program was unsuccessful in assigning a valid imputed value.
- Any instances where the imputed value was not consistent with preexisting nonmissing values were noted.
- It was necessary to ensure that any restrictions on the final imputed value for a given nonrespondent were honored. For example, some respondents were known to be employed, but either full-time or part-time employment status was not known. Checks were conducted to ensure that these respondents had either full-time or part-time status (but not unemployed or other status) assigned to the employment status variable IRWRKSTAT.

³ Refer to Appendix E for more details about likeness constraints.

- When multiple missingness patterns existed or different restrictions were placed on final imputed values, each pattern was treated separately, and the distribution of imputed values within each missingness pattern, or restricted imputation class, was investigated. For example, if it was known that a respondent was a past year user, both past month and past year users should have been included among the imputed values, not just past month users.

Appendix G: Household Roster Edit Checks

G.1 Introduction

In the household roster for the 2023 National Survey on Drug Use and Health (NSDUH), the interviewer or respondent was asked to enter a roster of the respondent's entire household, which included age, gender, and the relationship to the respondent. This appendix details the edits that were applied to check for, and if possible, resolve illogical ages, genders, or relationship codes in the household roster.

These edits can be divided into two categories: (1) checks within the respondent's roster alone, and (2) checks based on the respondent's roster and the other pair member's roster. These edits are detailed in Sections G.2 and G.3. Although the edits listed in Section G.2 were flagged using only the respondent's roster, when available, the other pair member's roster and screener roster were reviewed before the final edits were applied.

G.2 Checks within the Respondent's Roster

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent. All responses flagged by these edits underwent manual review. [Table G.1](#) lists the number of times each of these edits was applied and whether the automated resolution, if any, was retained or whether another edit was applied during review. For many of these edits, the automated resolution is to set the relationship code to bad data. This is frequently changed to a valid code during review using information obtained from the other pair member or the screener. Of the 1,190 records with a relationship code automatically set to bad data based on these edits in 2023, 544 remained as bad data after editing was complete.

In 2023, there were 830 respondents with rosters that were flagged by one or more of the edits listed in [Table G.1](#). Often these rosters were flagged for multiple reasons, or even multiple times for the same reason. For example, a respondent may list two children-in-law but no children. In that situation, both children-in-law in the roster would be flagged for the same reason. The exception to this is that edits 1 through 20 were applied first and reviewed together. The remaining edits were applied afterward, and responses that had already been reviewed were not flagged a second time. The counts in [Table G.1](#) are at the roster level rather than the respondent level.

Table G.1 Checks within the Respondent's Roster

					Final Resolution		
#	Description		Internal Check ¹	Automated Resolution	Flagged Occurrences	Keep Automated Resolution	Other Edit
1	Respondent lists multiple spouses, ex-spouses, or multiple unmarried partners.		1	Set all spouses ² to bad data.	177	120	24
2	Respondent lists both a spouse and an ex-spouse or an unmarried partner.		1	Retain spouse and set ex-spouse or partner to bad data.	22	6	2
3	Parent is younger than respondent.	Parent is younger than 15.	2	Set parent to bad data.	112	66	42
		Otherwise	2	No change. Output for review.	32	N/A	16
4	Child is older than respondent.	Respondent is younger than 15.	3	Set child to bad data.	91	49	42
		Otherwise	3	No change. Output for review.	135	N/A	45
5	Parent is less than 12 years older than respondent.		-	Set parent to bad data.	63	28	16
6	Mother is more than 60 years older than respondent.		-	Set mother to bad data.	10	5	3
7	Child is less than 12 years younger than respondent and respondent's spouse. ²		-	Set child to bad data.	301	173	118
8	Sibling is older than youngest parent.		-	Set sibling to bad data.	56	26	9
9	Parent is younger than the oldest sibling.		-	Set parent to bad data.	18	6	6
10	Child-in-law is older than respondent.	Child-in-law is at least 10 years older.	7	Set child-in-law to bad data.	4	0	4
		Otherwise	7	No change. Output for review.	3	N/A	1
11	Parent-in-law is younger than respondent.	Parent-in-law is at least 10 years younger.	7	Set parent-in-law to bad data.	4	3	1
		Otherwise	7	No change. Output for review.	2	N/A	1

Table G.1 Checks within the Respondent's Roster (continued)

#	Description	Internal Check ¹	Automated Resolution	Flagged Occurrences	Final Resolution	
					Keep Automated Resolution	Other Edit
12	Respondent lists a parent- or child-in-law, but either the respondent or the roster member is younger than 15.	6	Set parent-/child-in-law to bad data.	88	55	33
	Otherwise	6	No change. Output for review.	23	N/A	9
13	Respondent lists two or more children-in-law but no children.	-	Set children-in-law to bad data.	6	6	0
14	Respondent lists two or more siblings-in-law but no spouse, and the respondent or sibling-in-law is 18 or younger.	-	Set sibling-in-law to bad data.	72	45	23
	Otherwise	-	No change. Output for review.	19	N/A	0
15	Respondent lists grandchild and either grandchild is older than respondent, or either respondent or respondent's spouse is younger than 26.	9	Set grandchild to bad data.	2	2	0
	Otherwise	9	No change. Output for review.	1	N/A	1
16	Respondent lists parent(s) and grandchild but no children and is less than 24 years older than grandchild.	3	Set grandchildren to bad data.	1	1	0
17	Sibling is more than 20 years older than respondent but less than 4 years younger than oldest parent.	-	Set sibling to bad data.	8	2	5
18	Respondent lists a grandparent or grandchild with less than a 20-year age difference.	2,3,9	Set relationship to bad data.	2	2	0
19	Respondent lists more than two parents.	-	Set all parents to bad data.	167	87	15

Table G.1 Checks within the Respondent's Roster (continued)

#	Description	Internal Check ¹	Automated Resolution	Flagged Occurrences	Final Resolution	
					Keep Automated Resolution	Other Edit
20	Respondent is at least 20 years older than a grandparent or 20 years younger than a grandchild.	-	Change from grandparent to grandchild or vice versa.	9	9	0
21	Respondent has a parent, no spouse, and lists a child who is less than 13 years younger than the respondent.	-	Change child to sibling.	3	1	0
22	Respondent has a grandparent, is 25 or younger, and lists a child or child-in-law who is less than 13 years younger than the respondent.	No nonimmediate family members ² in the household.	Change child(-in-law) to parent.	1	1	0
	Otherwise	-	Set child(-in-law) to bad data.	0	0	0
23	Respondent lists only one parent and a spouse ³ who is at least 15 years older than the respondent, and respondent is either younger than 17 or between 17 and 20 and the spouse is older than the parent.	4	Change spouse to parent.	5	4	1
24	Respondent lists a spouse, and either respondent or spouse is younger than 15.	4,5	Change spouse to bad data.	12	2	4
25	Sibling is at least 20 years younger than respondent, respondent is older than 25 and lists no parents, or mother is more than 50 years older than sibling.	-	Change sibling to child.	74	54	6
26	Sibling is at least 25 years older or younger than respondent, and respondent is part of a pair.	-	None. Flag for review.	16	N/A	7

N/A = not applicable.

¹ Internal consistency checks used in computer-assisted interviewing are listed in Section 11.2.1.2.² Nonimmediate family codes are 16, 17, 18, or 19, as described in [Table 11.2](#).³ Includes partner and ex-spouse.

G.3 Checks Using the Respondent's and the Other Pair Member's Rosters

The following edits were performed on the roster relationship codes and ages, where the relationship code given was internally inconsistent, and information from the other pair member could be used to reconcile the inconsistency for cases where two household members were interviewed. As with the edits detailed in Section G.2, all responses flagged by these edits

underwent manual review. In 2023, there were 64 respondents with rosters that were flagged by one or more of the edits listed in [Table G.2](#). These edits were all applied after those listed in [Table G.1](#), and responses that had already been reviewed were not flagged a second time. The two respondents are referred to as pair member “A” and pair member “B.” All of these checks were performed for each pair member. For brevity, only those checks where pair member “A’s” roster was changed are listed in [Table G.2](#); however, the counts include the checks for both pair member “A” and pair member “B.” As with [Table G.1](#), the counts in [Table G.2](#) are at the roster level.

Table G.2 Checks across Pair Members’ Rosters

#	Edit Failure	Automated Resolution	Flagged Occurrences	Final Resolution	
				Keep Automated Resolution	Other Edit
1	Both respondents are 18 or younger and list at least one parent. Pair member “A” listed a child younger than 25 that was set to bad and listed fewer siblings than pair member “B.” ¹	Change from bad data to sibling.	25	0	25
2	Both respondents list spouses, ² at least one is older than 21, and they are at least 10 years apart. “A” is younger than 26, lists only one parent, is at least 3 years older or younger than “B’s” spouse, and lists a spouse who is at least 15 years older. “B” lists no parents. ³	Change “A’s” spouse ¹ to a parent.	0	0	0
3	Both respondents list spouses ² and are younger than 21. “A” lists only one parent, and spouse is at least 15 years older than “A.”	Change “A’s” spouse ¹ to a parent.	0	0	0
4	“A” lists a spouse and no parents-in-law and (1) “A” is younger than 15, (2) spouse is younger than 15 and “B” does not list a spouse, or (3) “A” is younger than 18 and said s/he was never married.	“A” lists one parent and “B” lists two.	Change “A’s” spouse to a parent.	0	0
		“B” lists one more sibling than “A.”	Change “A’s” spouse to a sibling.	1	0
		Either “A” or “A’s” spouse are 16 or younger.	Change “A’s” spouse to bad data.	0	0
		“A” lists no parents and “B” lists no spouses or partners.	Change “A’s” spouse to bad data. If spouse is younger than 15, change spouse age to bad data.	0	0
		Otherwise	No change. Output for review.	1	N/A
5	“A” lists a partner or ex-spouse and is either younger than 15 or the partner is younger than 15, and “B” does not list a partner.	“A” lists one parent and “B” lists two.	Change “A’s” partner to a parent.	0	0
		“B” lists one more sibling than “A.”	Change “A’s” partner to a sibling.	0	0
		Otherwise	No change. Output for review.	0	N/A

Table G.2 Checks across Pair Members' Rosters (continued)

#	Edit Failure	Automated Resolution	Flagged Occurrences	Final Resolution	
				Keep Automated Resolution	Other Edit
6	Both "A" and "B" list a spouse, ² "A's" spouse is the same age and gender as "B's" spouse, and "A" lists at least one sibling.	Change "A's" spouse to bad data.	2	0	0
7	"A" is younger than 26 and lists a grandchild.	"B" is also younger than 26, lists at least one sibling, and lists parents or grandparents of the same age as "A."	0	0	0
	"A" and "B" have no bad relationship codes on their rosters, and "B" is not "A's" oldest grandparent.	Change "A's" grandchild to "Other Relative."	0	0	0
	Otherwise	Change "A's" grandchild to bad data.	0	0	0
8	"A" is younger than 26, lists a child or child-in-law who is less than 13 years younger than "A," and lists at least one grandparent. "B" is also younger than 26, and one of "B's" parents could be "A's" grandparent.	Change "A's" child or child-in-law to "Other Relative."	0	0	0
9	"A" lists a household member aged <i>N</i> and has fewer household members aged <i>NN</i> than "B," or vice versa, where <i>N</i> is an integer between 1 and 9 (e.g., 1 vs. 11 or 33 vs. 3).	No change. Output for review.	340	N/A	4

N/A = not applicable.

¹ The two respondents are referred to as pair member "A" and pair member "B."

² Includes partner and ex-spouse.

³ Includes parents-in-law.

Appendix H: Rules for Determining Pair Relationships

H.1 Rules for Determining Matching Pairs, in Priority Order

The following rules, called match types, are used to determine the member of a respondent's household roster that corresponds to the other pair member. In these rules, an "age match" occurs if the questionnaire age of one pair member matches a roster age in the other pair member's roster, and a "gender match" occurs if the questionnaire gender of one of the pair members matches a roster gender in the other pair member's roster. The rules are listed in priority order in [Table H.1](#), along with the number of pairs to which each rule was applied. In the table below, if the rules for Pair Member A and Pair Member B in a single row differ (e.g., match type = 1.00), then the count for that row includes the rules as listed and the rules with Pair Member A and Pair Member B reversed. If the age and/or gender are off when finding these matches, the age and/or gender are defined by the questionnaire age and gender of the selected pair member when determining the pair domain. It was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either 0 or 1 MBRSEL for each responding pair. Codes for instances where MBRSEL was greater than 1 were removed and the match type index was adjusted accordingly.

Table H.1 Rules for Determining Matching Pairs, in Priority Order

Match Type	Rule		Count
	Pair Member A	Pair Member B	
0.00	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	9,915
1.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	11,468
1.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	79
2.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	219
2.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	18
2.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	6
3.00	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL off or missing for all roster members	563
3.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL off or missing for all roster members	44
3.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL off or missing for all roster members	10
3.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age and gender match exactly, MBRSEL off or missing for all roster members	50
4.00	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	133

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
4.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	11
4.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
4.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age matches exactly, gender off, exactly one MBRSEL in right place	8
4.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	3
5.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	61
5.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	12
5.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	1
5.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	4
5.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	1
5.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	0
6.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	185
6.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	21
6.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	7
6.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	12
6.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	3
6.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	2

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
6.06	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	0
7.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	11
7.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	2
7.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	0
7.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	4
7.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	0
7.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes situations where one of closer age could have been selected	0
8.00	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	159
8.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	13
8.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Everything missing	1
8.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Everything missing	47
8.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Everything missing	1
8.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Everything missing	3
9.00	Age and gender match exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, only one possible match	1
9.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, only one possible match	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
9.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, only one possible match	0
9.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Gender and reported household sizes match exactly, age missing, only one possible match	0
9.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, only one possible match	0
9.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Gender and reported household sizes match exactly, age missing, only one possible match	0
10.00	Age and gender match exactly, exactly one MBRSEL in right place	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	3
10.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	1
10.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	1
10.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	3
10.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	0
10.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	1
10.06	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	0
10.10	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	0
11.00	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
11.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0
11.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0
11.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0
11.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0
11.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	0
12.00	Age and gender match exactly, exactly one MBRSEL in right place	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
12.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
12.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
12.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
12.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
12.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
12.10	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	Two matches on gender, age within one, MBRSEL off or missing for all roster members, both have same relationship code (randomly select one)	0
13.00	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL	20
13.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL	4
13.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL	0
13.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age within one, gender off, one MBRSEL	2
13.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL	1
13.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age within one, gender off, one MBRSEL	0
13.10	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	Age within one, gender off, one MBRSEL	0
13.11	Age and gender matched exactly for two or more roster members, with MBRSEL off or missing for all roster members; randomly select one with a matching relationship code	Age within one, gender off, one MBRSEL	0
13.13	Age within one, gender off, one MBRSEL	Age within one, gender off, one MBRSEL	0
14.00	Multiple matches or no match, but no relationship codes are missing, and none involve domains of interest	Anything	14
15.00	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL off or missing for all roster members	6
15.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL off or missing for all roster members	3
15.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL off or missing for all roster members	0
15.03	Age and gender match exactly, MBRSEL off or missing for all roster members	Age matches exactly, gender off, MBRSEL off or missing for all roster members	3
15.04	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL off or missing for all roster members	0
15.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	Age matches exactly, gender off, MBRSEL off or missing for all roster members	1
15.10	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	Age matches exactly, gender off, MBRSEL off or missing for all roster members	0

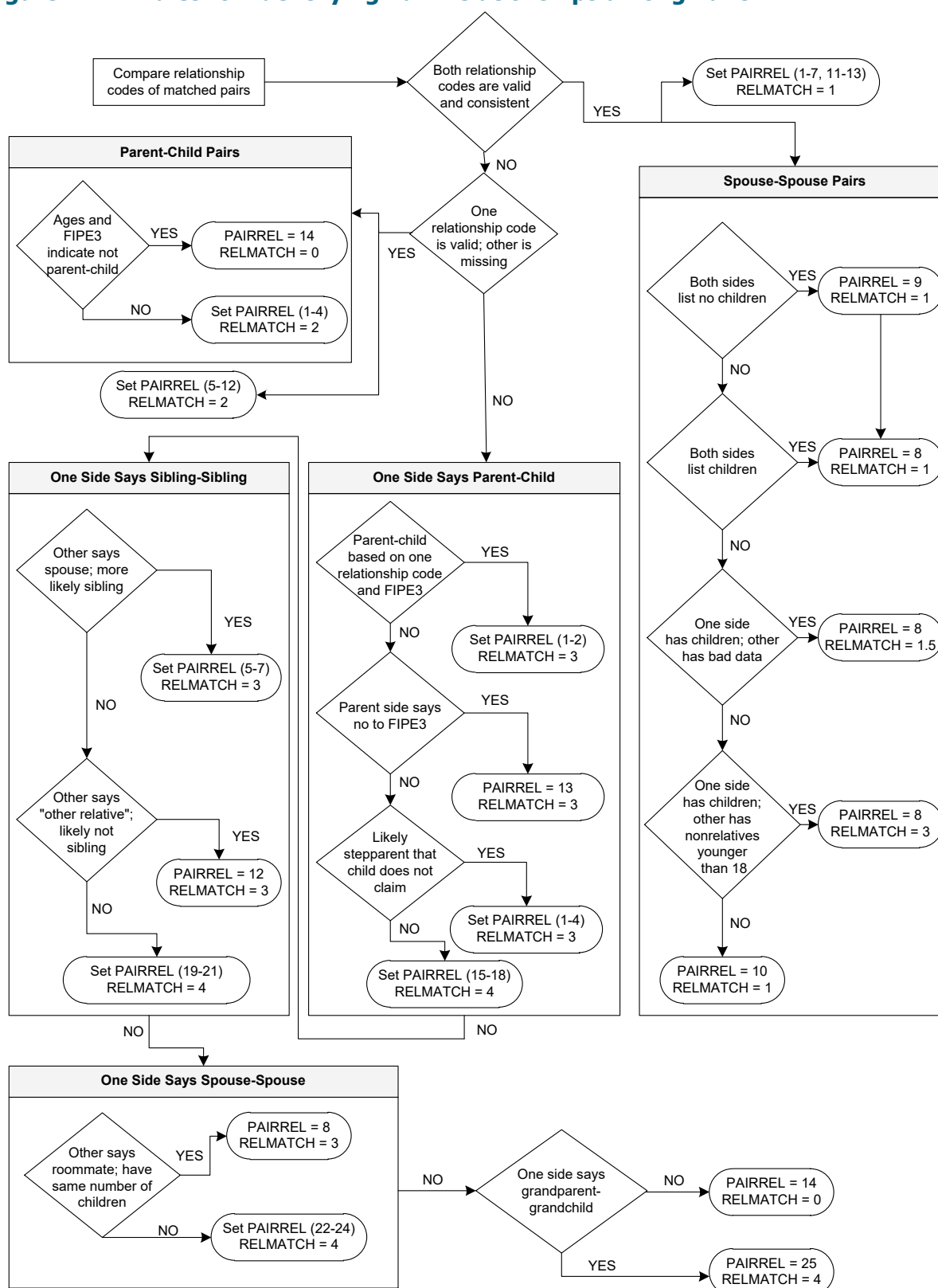
Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		Count
	Pair Member A	Pair Member B	
16.00	Age and gender match exactly, exactly one MBRSEL in right place	No match at all and involves domains of interest (often paired respondent is missing from roster)	137
16.01	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all and involves domains of interest (often paired respondent is missing from roster)	15
16.02	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all and involves domains of interest (often paired respondent is missing from roster)	4
16.03	Age and gender match exactly, MBRSEL off or missing for all roster members	No match at all and involves domains of interest (often paired respondent is missing from roster)	28
16.04	Age matches exactly, gender off, exactly one MBRSEL in right place	No match at all and involves domains of interest (often paired respondent is missing from roster)	1
16.05	Age within one, gender matches exactly, MBRSEL off or missing for all roster members	No match at all and involves domains of interest (often paired respondent is missing from roster)	5
16.06	Age within 10, gender matches, exactly one MBRSEL in right place, excludes situations where MBRSEL could have been applied to one of closer age	No match at all and involves domains of interest (often paired respondent is missing from roster)	0
16.10	Exactly 2 matches on age, gender, and relationship code; MBRSEL off or missing for all roster members; does not matter which match is picked (randomly select one)	No match at all and involves domains of interest (often paired respondent is missing from roster)	1
16.13	Age within one, gender off, one MBRSEL	No match at all and involves domains of interest (often paired respondent is missing from roster)	0
16.15	Age matches exactly, gender off, MBRSEL off or missing for all roster members	No match at all and involves domains of interest (often paired respondent is missing from roster)	0
17.00	No match at all and involves domains of interest	No match at all and involves domains of interest	24

H.2 Rules for Identifying Pair Relationships among Pairs

[Figure H.1](#) summarizes the rules used to identify the pair relationship, using the relationship codes and questionnaire ages of the two pair members. The variable PAIRREL identifies the pair relationship as defined by [Table 14.3](#) in the main body of this report. In this flowchart, “children” are defined as roster members listed as children with nonmissing ages less than 18. The question FIPE3 asks whether the respondent is the parent of a selected 12- to 17-year-old. The variable RELMATCH indicates the quality of the match between pair members, as defined in [Table 14.4](#) in the main body of this report.

Figure H.1 Rules for Identifying Pair Relationships among Pairs



Appendix I: Conditions Used to Reconcile Differing Multiplicity Counts between Pair Members

I.1 Introduction

To determine multiplicity counts, counts were obtained from each pair member. The count from the pair member who was the focus member of the domain was considered the direct count, and the count from the other pair member was considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided neither roster had bad data codes. The strategy was usually clear if one pair member had bad data in the household roster or had a zero count when the pair relationship precluded a value of zero. The count from the pair member with good, nonzero data was usually preferred in those situations. If the bad data were limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes gave a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts disagreed. The rules that were used to reconcile these disagreeing counts are outlined in this appendix.

Note that the reconciliation of differing counts was necessary for parent-child and sibling-sibling pairs but not for spouse-spouse pairs, because the multiplicity count for spouse-spouse pairs was always one. As noted in Section 14.3.1, it was technically possible for a respondent to have multiple spouses, but these situations were not accounted for in processing.

I.2 Parent-Child Counts

For parent-child counts, the screener was used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus.

Parent-child pairs, child focus: The multiplicity counts in this domain reflect the selected child's parents and most had values of 1 or 2. [Figure I.1](#) illustrates the rules that were applied to determine the multiplicity counts in this domain.

Parent-child pairs, parent focus: The multiplicity counts in this domain reflected the selected parent's children and were limited to have values of at least 1. [Figure I.2](#) illustrates the rules that were applied to determine the multiplicity counts in this domain.

Figure I.1 Rules for Determining the Multiplicity Count: Parent-Child Pairs, Child Focus

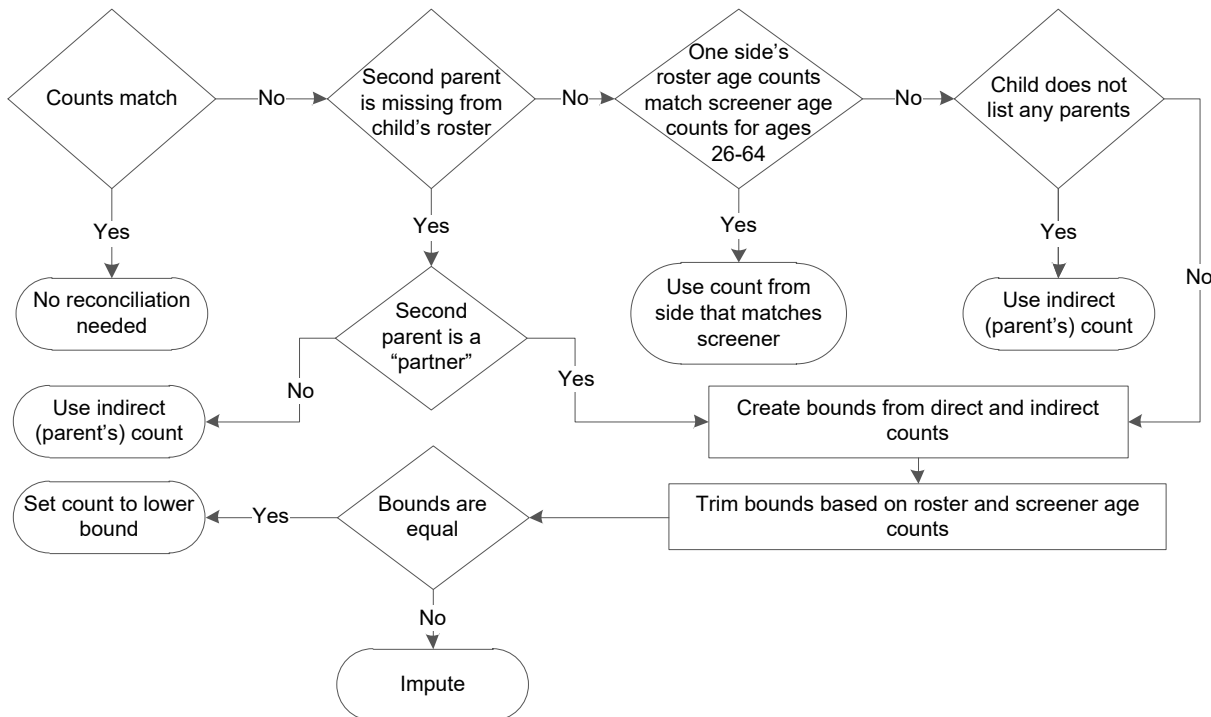
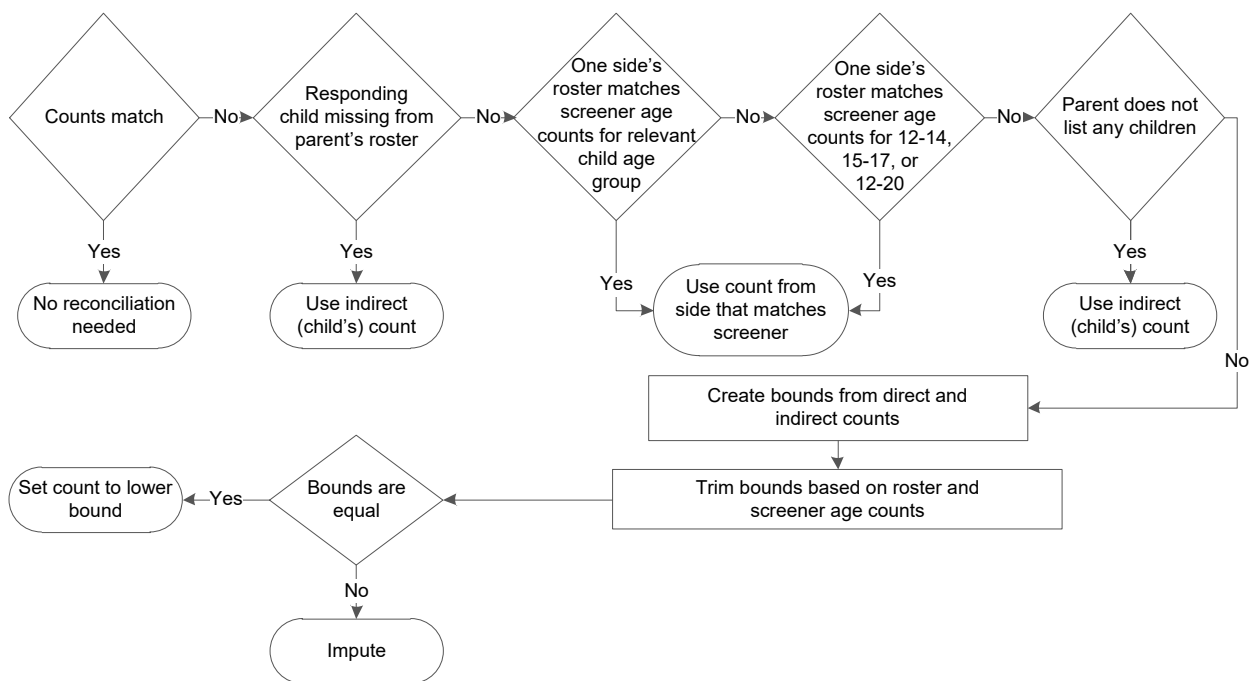


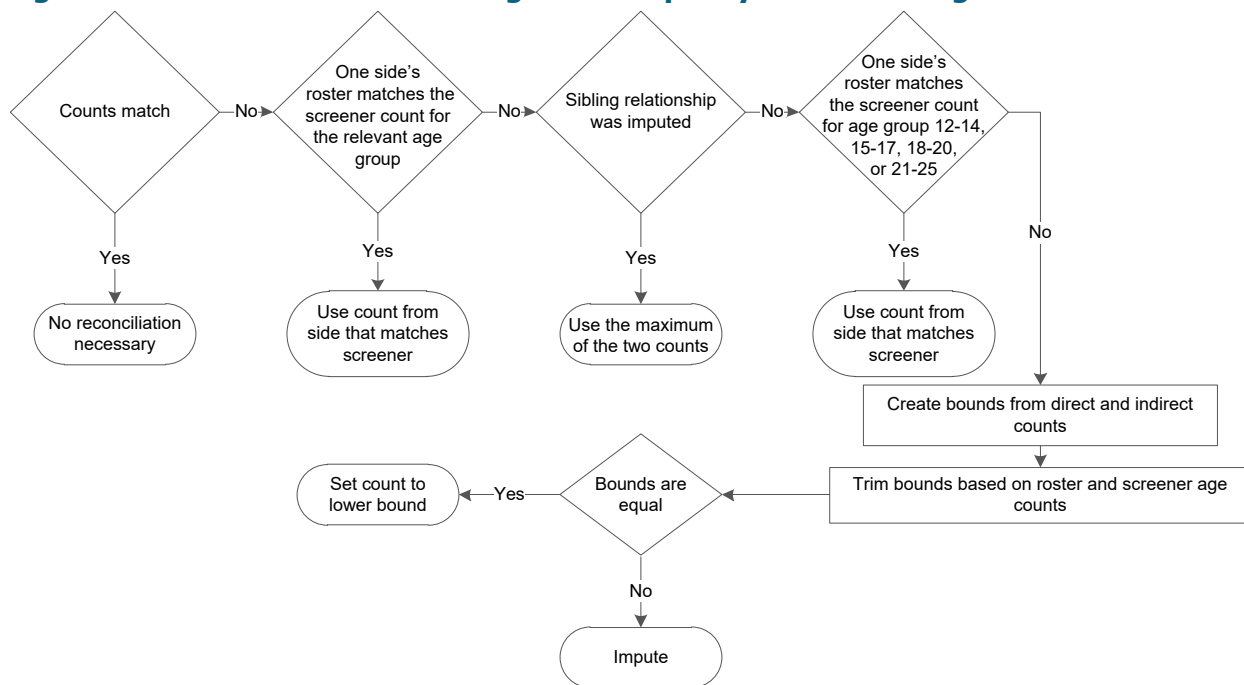
Figure I.2 Rules for Determining the Multiplicity Count: Parent-Child Pairs, Parent Focus



I.3 Sibling-Sibling Counts

Although there were two types of sibling-sibling pairs under consideration, each associated with two domains, the same rules were applied to all four domains. When the older sibling was the focus, the multiplicity count was a count of the number of siblings within the younger age group (12 to 14 or 12 to 17). Conversely, the multiplicity count was the number of siblings in the older age group (15 to 17 or 18 to 25) when the younger sibling was the focus. [Figure I.3](#) illustrates the rules that were applied to determine the multiplicity counts for these domains.

Figure I.3 Rules for Determining the Multiplicity Count: Sibling Pairs



Appendix J: Conditions Used to Reconcile Differing Household-Level Person Counts between Pair Members

J.1 Introduction

This appendix discusses the conditions used to reconcile household-level person counts between pair members. Typically, the household-level person counts between the two pair members were in agreement, and the determination of the final household-level count was straightforward. The rules that were used to assign a final count and reconcile disagreeing counts are outlined in this appendix. If the final count could not be determined based on the pair members' rosters, then upper and lower bounds were created and the final count was left to imputation.

J.2 Parent-Child Counts

The process for reconciling parent-child counts across pairs differed based on the member of focus and whether there was more than one family unit in the household. [Figures J.1](#) and [J.2](#) outline these processes for the child-focus and parent-focus counts, respectively. The figures refer to the pair members as sides A and B, and most conditions are written assuming side A is the trusted side. This is for simplicity of presentation. During processing, each condition was applied to both sides. Additionally, only the process for the 12-20 age group is shown. The few variations in this process based on the age group are noted below each figure.

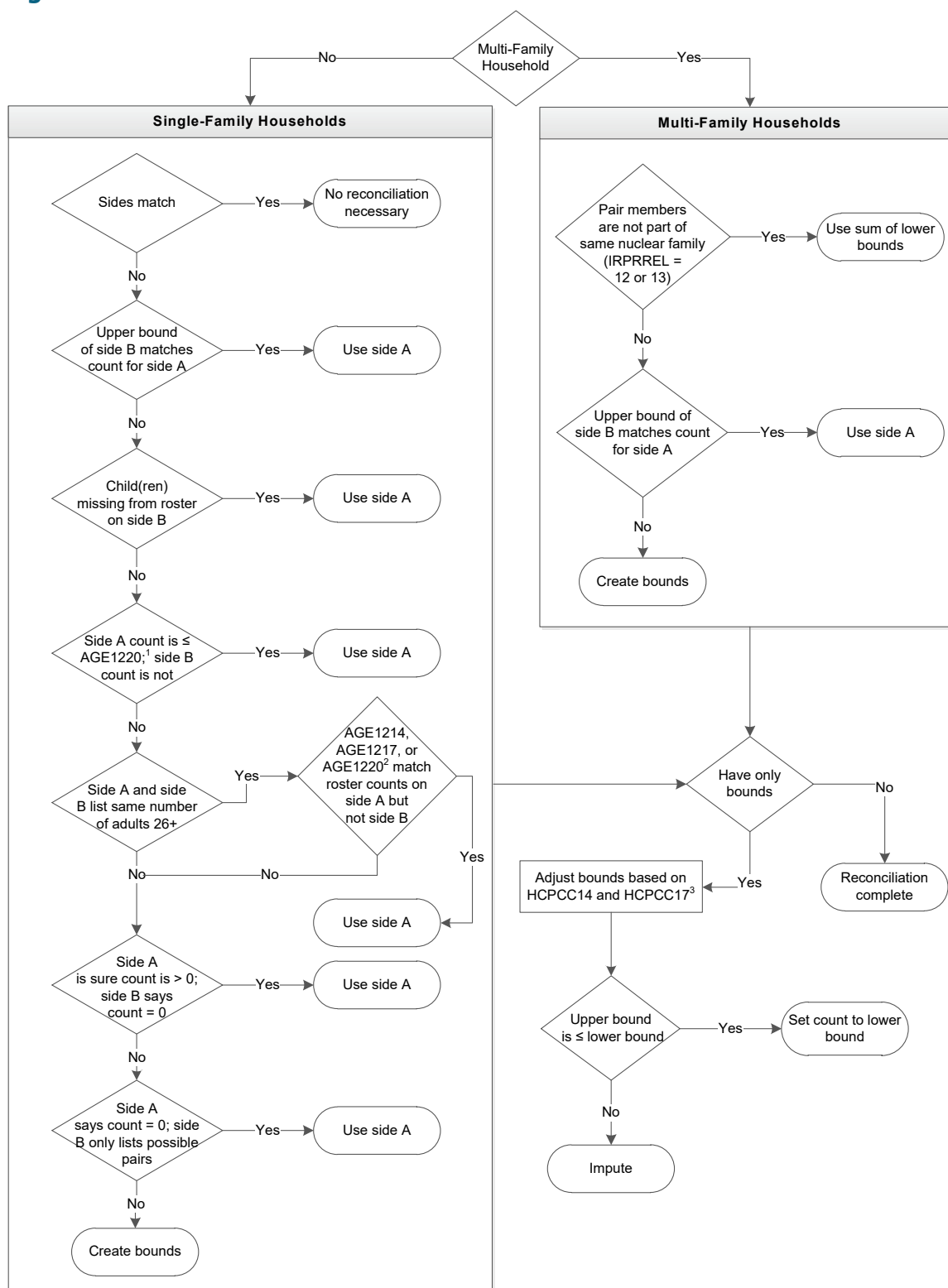
J.3 Sibling-Sibling Counts

[Figure J.3](#) outlines the process for reconciling the household sibling counts between pairs for the 12-14 and 15-17 age groups. The process for the 12-17 and 18-25 age groups was similar.

J.4 Spouse-Spouse Counts (with or without Children)

[Figure J.4](#) outlines the process for reconciling the spouse-spouse counts between pairs. After reconciling the spouse-spouse counts, the conditions listed in [Figure J.5](#) were used to reconcile the counts of spouse-spouse pairs with children younger than 18.

Figure J.1 Reconciliation of Parent-Child 12-20 Household Counts: Child Focus

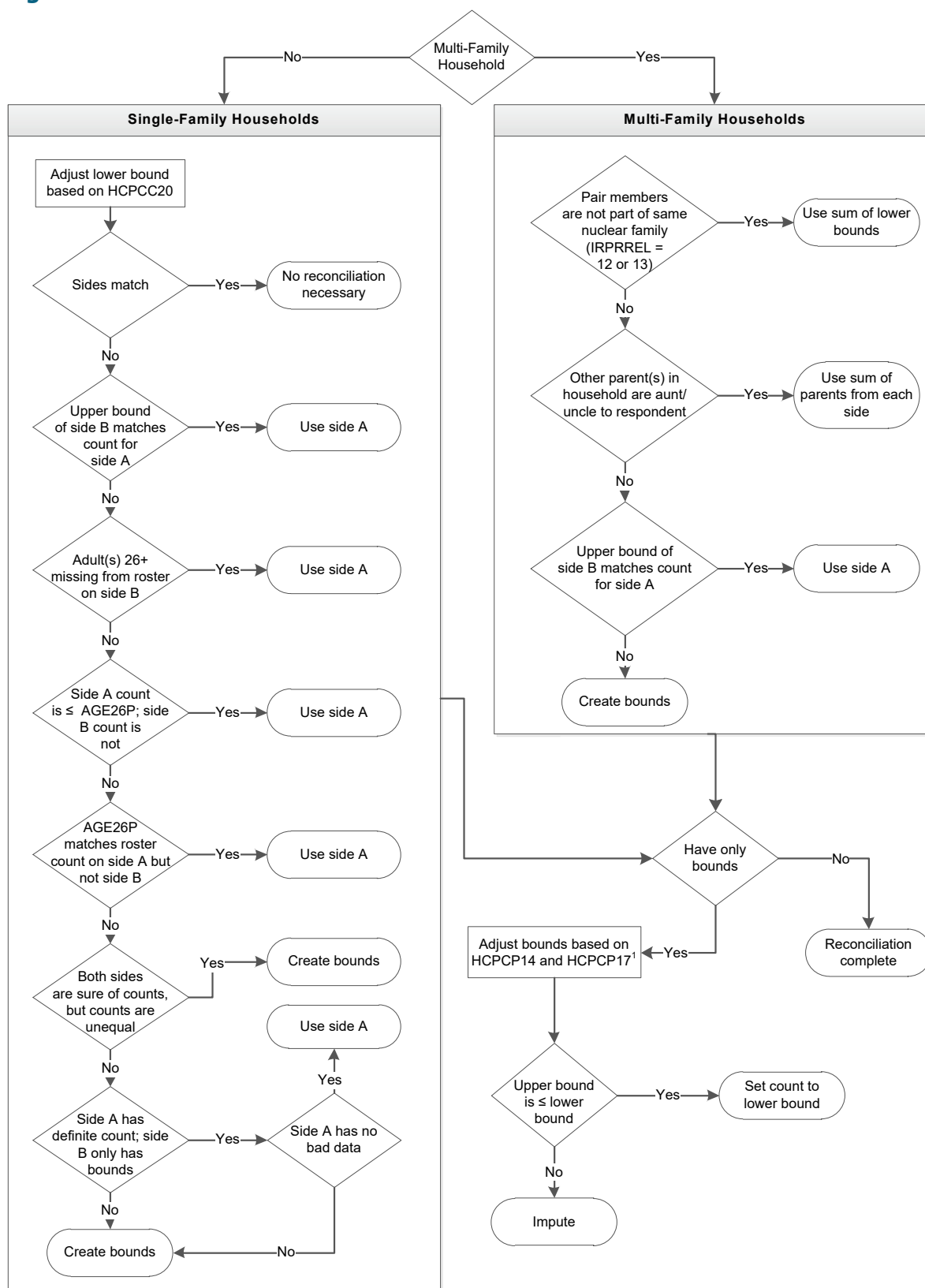


¹ AGE1214 and AGE1217 were used for the 12-14 and 12-17 age groups, respectively.

² For the 12-14 age group, only AGE1214 was used. For the 12-17 age group, AGE1214 and AGE1217 were used.

³ Only HCPCC14 was used for the 12-17 age group, and this step was skipped entirely for the 12-14 age group.

Figure J.2 Reconciliation of Parent-Child 12-20 Household Counts: Parent Focus



¹ Only HCPCP14 was used for the 12-17 age group, and this step was skipped entirely for the 12-14 age group.

Figure J.3 Reconciliation of Sibling-Sibling 12-14 and 15-17 Household Counts

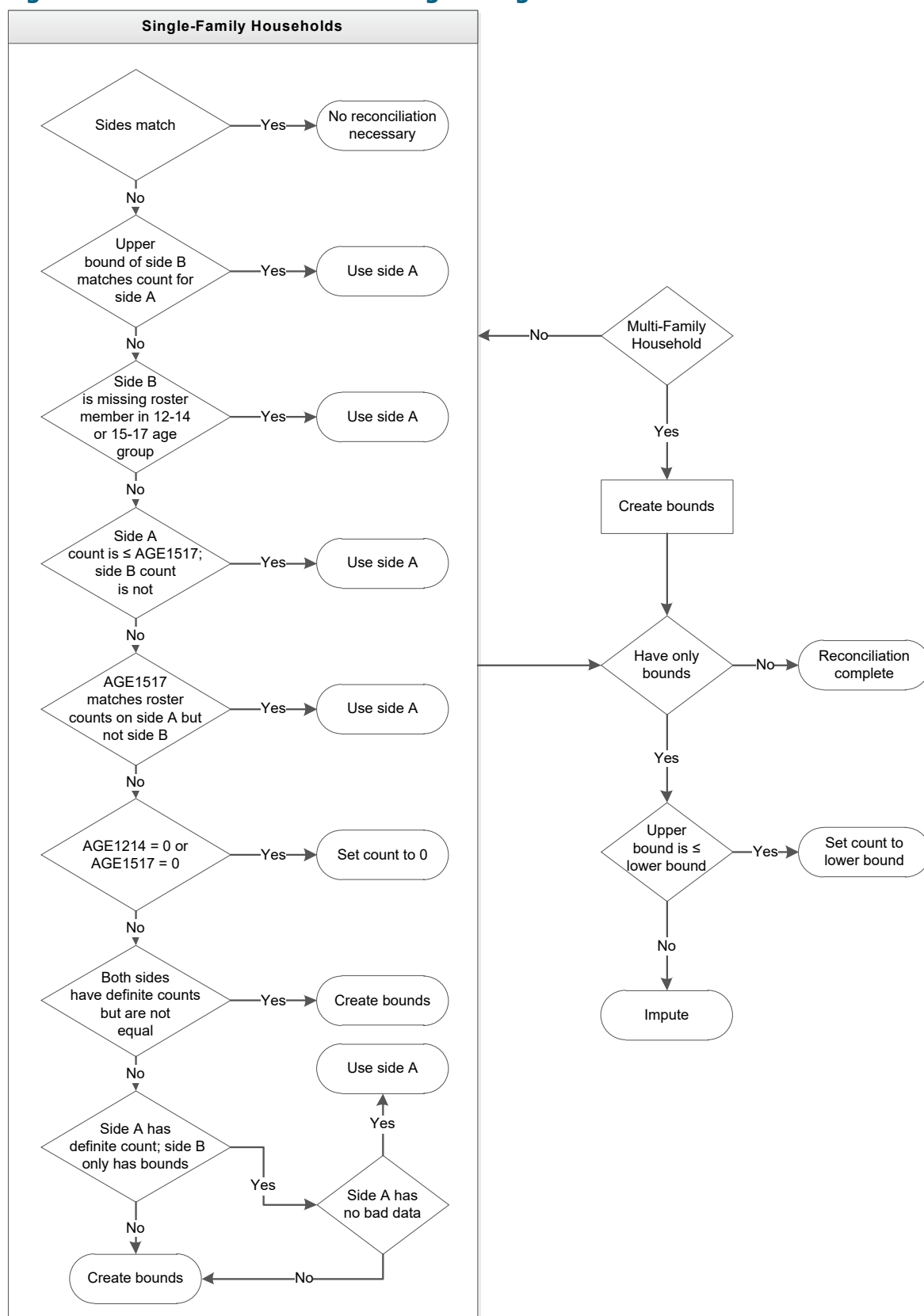


Figure J.4 Reconciliation of Spouse-Spouse Household Counts

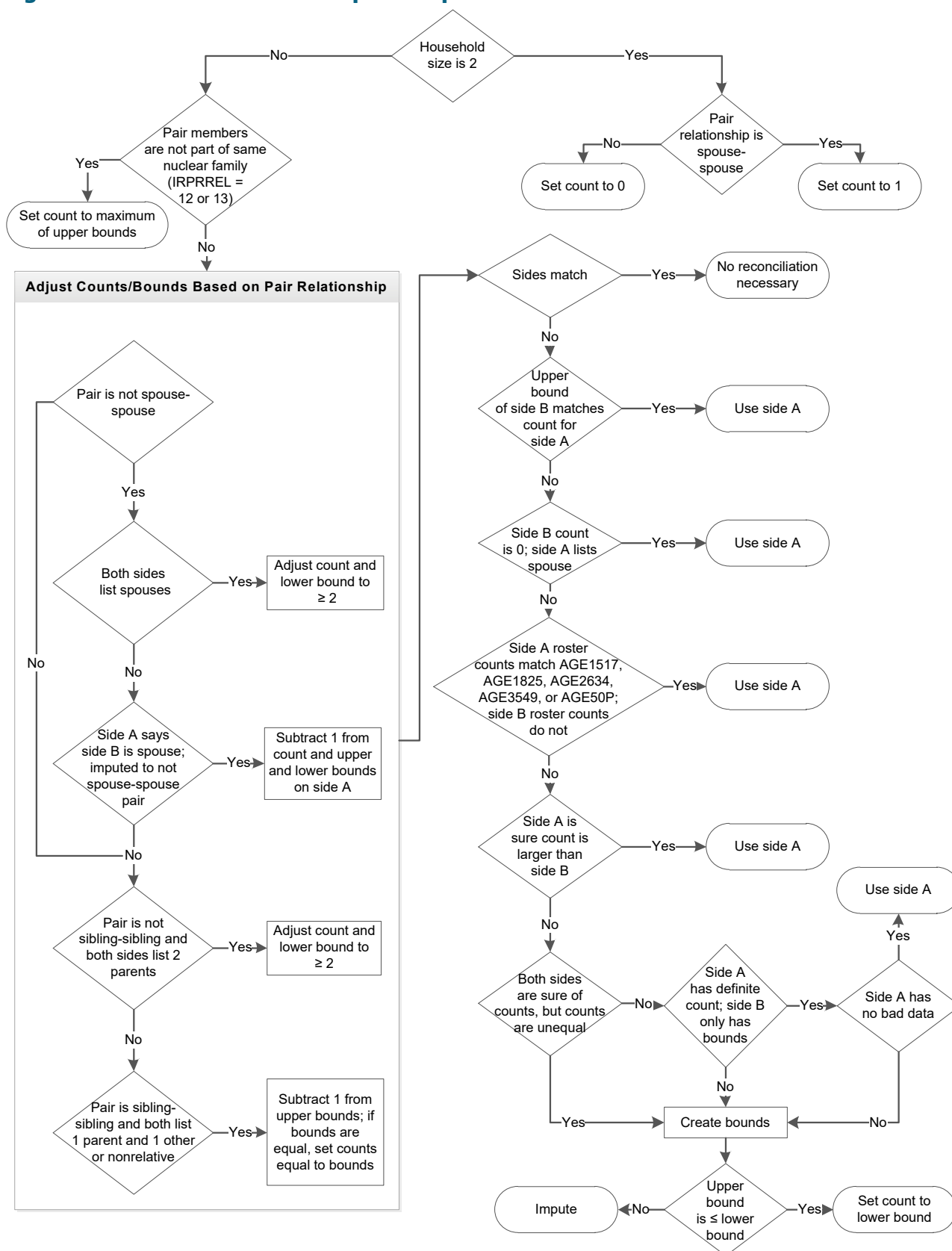
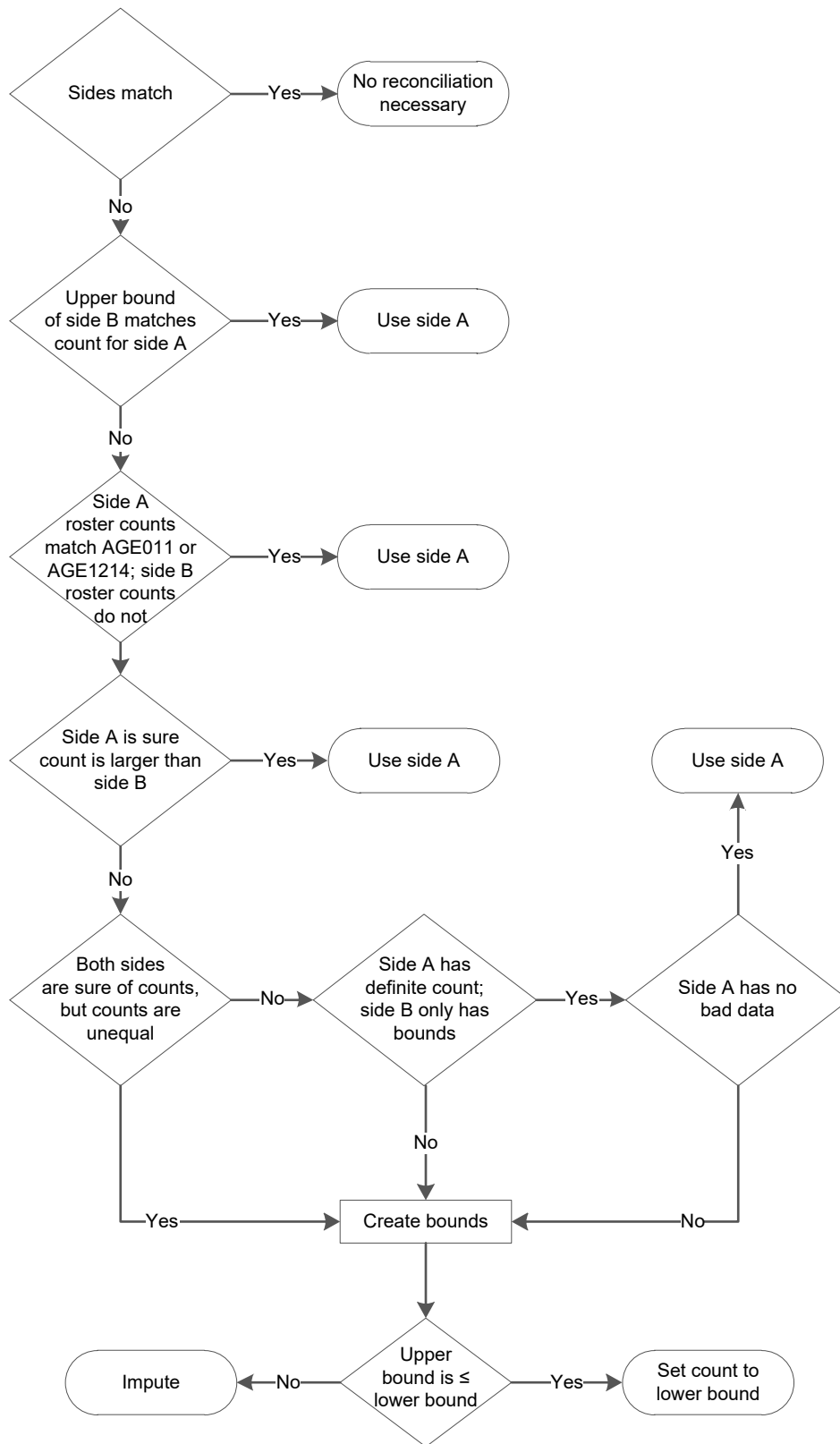


Figure J.5 Reconciliation of Spouse-Spouse with Children Younger than 18 Household Counts





Released 2025

SAMHSA's mission is to lead public health and service delivery efforts that promote mental health, prevent substance misuse, and provide treatments and supports to foster recovery while ensuring equitable access and better outcomes.

1-877-SAMHSA-7 (1-877-726-4727) | 1-800-487-4889 (TDD) | www.samhsa.gov