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

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

A.1 Distance Function

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

$$\Delta(w,d) = \sum_{k \in s} \frac{d_k}{A_k} \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\} \tag{A1.1}$$

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

$$\sum_{k \in s} x_k d_k a_k = T_x \tag{A1.2}$$

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

$$a_k(\lambda) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp\{ A_k x_k' \lambda \}}{(u_k - c_k) + (c_k - \ell_k) \exp\{ A_k x_k' \lambda \}} \tag{A1.3}$$

Note that the number of parameters in GEM should be $\leq n$, where n is the size of the sample s. This is also the dimension of vectors d and w. It follows from Equation A1.3 that

$$\ell_k < a_k < u_k,\ k = 1, \ldots, n \tag{A1.4}$$

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

$$\Delta(w,d) = \sum_{k \in s} d_k a_k \log a_k - \sum_{k \in s} d_k (a_k - 1) \tag{A1.5}$$

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

The logit method of Deville and Särndal (1992) is also a special case of GEM by setting $\ell_k = \ell, u_k = u,$ and $c_k = 1$ for all k.

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

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (A1.3) can be justified for all three types of adjustment. Denote the winsorized weights by $\{b_k\}$ where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = med\{d_k\} \pm 3 *$ IQR (where IQR denotes the interquartile range) if d_k is an extreme weight (where the quartiles for the weights are defined with respect to a suitable design-based stratum).

For the nonresponse adjustment, the sample is first divided into two parts: s^*, the non-extreme weight subsample; and s^{**}, the extreme weight subsample. For non-extreme weights, the following are set: $\ell_2 = 1, c_2 = \rho^{-1}, u_2 = u > \rho^{-1}$, where ρ is the overall response propensity; and for extreme weights with high weights, they are $\ell_k = \ell m_k, c_k = \rho^{-1} m_k, u_k = u_1 m_k$, where $m_k = b_k / d_k$ and $1 \le \ell_1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, for extreme weights with low weights, $\ell_k = \ell_3 m_k, c_k = \rho^{-1} m_k, u_k = u_3 m_k$, and $1 \le \ell_3 < \rho^{-1} = c_3 < u_3$.

For the poststratification adjustment, for non-extreme weights, $\ell_k = \ell_2$, $c_k = c_2 = 1, u_k = u_2$, and for high extreme weights, $\ell_2 = \ell_1 m_k, c_k = m_k, u_k = u_1 m_k$, and, similarly for low extreme weights, $\ell_k = \ell_3 m_k, c_k = m_k, u_k = u_3 m_k$. The extreme-value adjustment is identical to poststratifcation, except for tighter bounds on extreme weights resulting from the final poststratification.

Notice that GEM allows the flexibility of specifying different bounds for different subsamples; in addition, the lower bound (in the case of nonresponse adjustments) can be made to equal one by choosing the center $c_k > 1$.

A.3 Newton-Raphson Steps

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

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

where

$$\phi_k^{(v)} = \left[\left(u_k - a_k^{(v)}\right)\left(a_k^{(v)} - \ell_k\right)\right] \Big/ \left[\left(u_k - c_k\right)\ \left(c_k - \ell_k\right)\right]$$

then, for Newton-Raphson iteration v, the value of the p-vector λ is adjusted as

$$\gamma^{(v)} = \gamma^{(v-1)} + \left(X'\Gamma_{\phi,v-1}X\right)^{-1}\left(T_x - \hat{T}_x^{(v-1)}\right)$$

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

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

A.4 Scaled Constrained Exponential Model

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

Appendix B: Derivation of Poststratification Control Totals

Appendix B: Derivation of Poststratification Control Totals

Unlike the person-level poststratification adjustment, the control totals for questionnaire dwelling unit (QDU)–level and person pair–level weight calibration cannot be derived from the U.S. Census directly. Estimates of the number of households and person pairs are not available at the domains we would like to control, and person-pair population estimates are not available even at a national level. However, by taking advantage of the two-phase design of the National Survey on Drug Use and Health (NSDUH), the screener dwelling unit (SDU) sample weights can be poststratified to U.S. Census population estimates. The calibrated SDU weights then can be used as stable control totals for the QDU- and person pair–level sample weights. In addition to the SDU weights, the person pair–level weights are calibrated to a second set of controls derived from the questionnaire, called household-level person counts. These controls are applied to pairs that are members of the 10 selected pair domains given below.

1. Parent-child pairs, child aged 12 to 14, target population is parents whose children aged 12 to 14 live with them;
2. Parent-child pairs, child aged 12 to 14, target population is children aged 12 to 14 living with their parents;
3. Parent-child pairs, child aged 12 to 17, target population is parents whose children aged 12 to 17 live with them;
4. Parent-child pairs, child aged 12 to 17, target population is children aged 12 to 17 living with their parents;
5. Parent-child pairs, child aged 12 to 20, target population is parents whose children aged 12 to 20 live with them;
6. Parent-child pairs, child aged 12 to 20, target population is children aged 12 to 20 living with their parents;
7. Sibling-sibling pairs, older sibling aged 15 to 17, younger sibling aged 12 to 14, target population is siblings aged 15 to 17 whose siblings are aged 12 to 14;
8. Sibling-sibling pairs, older sibling aged 18 to 25, younger sibling aged 12 to 17, target population is siblings aged 18 to 25 whose siblings are aged 12 to 17;
9. Spouse-spouse and partner-partner pairs; and
10. Spouse-spouse and partner-partner pairs with children under the age of 18 living in the household.

B.1 Derivation of QDU–Level Poststratification Controls

The derivation of QDU-level poststratification controls is not directly possible. Instead it must be based on work done for the person-level calibration. At the person level, weights are calibrated to the control totals we wish to reach. These weights are then altered in order to conform to use with QDU-level data.

B.1.1 Person Level

B.1.1.1 Receiving and Deriving Person-Level Poststratification Control Totals

Civilian, noninstitutionalized population estimates for ages 12 and older are provided by the Population Estimates Branch of the U.S. Bureau of the Census. We receive two files, one at the national level and the other at the State level, each containing estimates of the population broken down by levels of month (12 levels), Hispanicity (2), race (6), sex (2), and age (6).

The breakdown received from the Census does not match the levels of the domains we would like to control. To account for this, we collapse levels. From this altered data, we create data sets with model group specific control totals. Observations in these data sets correspond to a breakdown by quarter (4), Hispanicity (2), race (5), sex (2), age (6), and number of States[16] in the model group (number of States varies according to which Census Region is represented in the model group).

B.1.1.2 Adjusting Screener Dwelling Unit Data to the Control Totals

In the person-level weighting, the SDU weights are poststratified to meet control totals based on the population estimates received from the Census. For NSDUH weighting, GEM is utilized to calibrate sample weights to multiple control totals. In doing so, each SDU receives an adjustment factor which, when multiplied by the initial weight, produces a final weight. The sum of all final weights corresponds to the civilian, noninstitutionalized population estimate for ages 12 and older, and the sum of all final weights in a domain corresponds to the control total for that domain. Note that there are a number of controls being calibrated to for each SDU, depending upon the domains to which the SDU belongs. The adjusted SDU weight reflects the civilian, noninstitutionalized population estimates for ages 12 and older, and can be utilized as a basis for constructing controls at the QDU and person pair levels.

B.1.2 QDU Level

B.1.2.1 Deriving QDU-Level Poststratification Control Totals from Adjusted SDU Weights

Since there are no controls for QDU-level poststratification available directly, we use the adjusted SDU weights. For these weights to be applicable at the QDU level, the SDU-level data must be restructured by sorting and summing over the domains to be used in the QDU-level calibration. This provides a data set where the summed weight, which still adds up to the proper population, is available for every domain to be utilized in the QDU calibration and, thus, can be used as a control total.

[16] The District of Columbia is included among States.

B.1.2.2 Adjusting QDU–Level Data to the Control Totals

As was done for the SDU data, the QDU-level data is adjusted via calibration in GEM of sample weights to multiple control totals. Each QDU receives an adjustment factor, similar to that described for the SDU weight in B1.1.2. The controls utilized in this calibration are based on the SDU weight as described in B.1.2.1 above. The adjusted weight is representative of the civilian, noninstitutionalized population estimates for ages 12 and older for all domains controlled within the modeling.

B.2 Derivation of Person Pair–Level Poststratification Controls

B.2.1 Deriving Person Pair–Level Poststratification Control Totals from Adjusted SDU Weights and the Household-Level Person Counts

Analogous to the QDU weights, some of the person pair controls are based on the SDU weights. However, two sets of control totals were utilized in the modeling, with one set based on the SDU weights and the other set based on the questionnaire roster.

For most pair data domains, those other than the 10 pair domains based on relationship, the control totals for the poststratification adjustments were obtained from SDU data, and were based on the number of possible pairs within SDUs. In order to obtain these pair counts belonging to various sociodemographic domains, the screener roster information was used to calculate all possible pairs within SDUs. For example, consider an SDU with two persons aged 12 to 17, and three persons aged 26 to 34. From this household composition, one can construct a single pair of persons aged 12 to 17, three pairs of persons aged 26 to 34, and six pairs of persons aged 12 to 17 and 26 to 34. It follows that the total number of possible pairs in this SDU is 10, from which the number of pairs belonging to the domain of interest can be obtained.

On the other hand, for the 10 selected pair domains based on relationship, the control totals for the poststratification adjustments were obtained from the questionnaire roster. This involved calibrating the pair weights to the number of persons in households belonging to each domain of interest. These controls were obtained from the larger sample of singles and pairs (i.e., one or two persons selected from DUs), and were calculated at the QDU (household) level. The pair weights were adjusted by the appropriate multiplicity. See Section 6.3 for details on the multiplicity counts and Section 6.4 on the household-level control totals, which are referred to there as household-level person counts.

B.2.2 Adjusting Person–Pair Level Data to the Control Totals

Like the SDU- and QDU-levels, the person pair-level data is adjusted via GEM. The use of two different types of controls requires a minor modification to the GEM macro so that both sets of controls may be addressed simultaneously. Similar to the SDU- and QDU-level poststratification steps, each pair receives an adjustment factor which, when multiplied by the initial weight, produces a final weight. The sum of all final weights corresponds to the civilian, noninstitutionalized population estimate for ages 12 and older, and the sum of all final weights in a domain corresponds to the control total for that domain.

Appendix C: GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

Appendix C: GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

Introduction

This appendix summarizes each questionnaire dwelling unit (QDU) model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this section provides a model-specific overview of weight calibration, as opposed to a State- or domain-specific one.

For 2003, modeling involved taking four model groups through three adjustment steps: (1) selected dwelling unit poststratification, (2) respondent dwelling unit nonresponse adjustment, and (3) respondent dwelling unit poststratification. After the final poststratification, the adjusted sampling weights were reasonably distributed and did not require the additional treatment of the ev step.

Model-specific summary statistics are shown in Tables C1a, C1b to C4a, and C4b. Included in these tables, for each stage of modeling, are: the number of factor effects included; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; weighted, unweighted and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. For more details on bounds, see Section 4.1. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the current weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data lead to situations whereby inclusion of all originally proposed levels of covariates in the model is not possible. The text and exhibits in Sections C1 to C4 summarize the decisions made with regard to final covariates included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit C.2; for the list of realized final model covariates, see Exhibits C1.1 to C4.3. The following sections establish a series of guidelines to assist in their interpretation.

C.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 3.1 (included here as Exhibit C.1 for easy reference) in Chapter 3. There, a complete list is provided of all variables and associated levels used at any stage of modeling. Note that not all factors or levels are present in all stages of modeling, and the initial set of variables is the same across model groups but changes for each stage of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment. Exhibits C1.1 to C4.3 provide lists of the proposed and realized covariates.

To help understand what effects are controlled for at each stage of the modeling, it may be useful to create cross-classification tables as shown in Section C.3. Sections C.2 and C.3 explain how to use various exhibits for selected model variables to construct these tables.

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables

Age [c]
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+ [1]

Gender [c]
1: Male, 2: Female [1]

Group Quarter Indicator [b]
1: College Dorm, 2: Other Group Quarter, 3: Nongroup Quarter [1]

Hispanicity [c]
1: Hispanic, 2: Non-Hispanic [1]

Household Size [c]
Continuous variable - count of individuals rostered with DU.

Household Type (ages of persons rostered within DU) [b]
1: 12-17, 18-25, 26+, 2: 12-17, 18-25, 3: 12-17, 26+, 4: 18-25, 26+, 5: 12-17, 6: 18-25, 7: 26+

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner) [b]
1: 50%-100% [1], 2: 10%->50%, 3: 0->10%

Percentage of Segments That Are Black (% black) [b]
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Percentage of Segments That Are Hispanic (% Hispanic) [b]
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Population Density [b]
1: MSA 1,000,000 or more, 2: MSA less than 1,000,000, 3: Non-MSA urban, 4: Non-MSA rural [1]

Quarter [bc]
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4 [1]

Race (3 levels) [c]
1: white, [1] 2: black, 3: Other

Race (5 levels) [c]
1: white, [1] 2: black, 3: Native American, 4: Asian, 5: multi-race.

Race of Householder [b]
1: Hispanic white [1], 2: Hispanic black, 3: Hispanic others, 4: Non-Hispanic white,
5: Non-Hispanic black, 6: Non-Hispanic others,

Relation to Householder [b]
1: Householder or Spouse, 2: Child, 3: Other Relative, 4: Nonrelative [1]

Segment Combined Median Rent and Housing Value (Rent/Housing) [b2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile [1]

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables (continued)

State [bc3]

Model Group 1: 1: Connecticut, 2: Maine, 3: Massachusetts, [1] 4: New Hampshire, 5: New Jersey, 6: New York, 7: Pennsylvania, 8: Rhode Island, 9: Vermont

Model Group 2: 1: Illinois, 2: Indiana, 3: Iowa, 4: Kansas, 5: Michigan, 6: Minnesota, 7: Missouri, 8: Nebraska, 9: North Dakota, 10: Ohio, 11: South Dakota, 12: Wisconsin [1]

Model Group 3: 1: Alabama, 2: Arkansas, 3: Delaware, 4: District of Columbia, 5: Florida, 6: Georgia, 7: Kentucky, 8: Louisiana, 9: Maryland, 10: Mississippi, 11: North Carolina, [1] 12: Oklahoma, 13: South Carolina, 14: Tennessee, 15: Texas, 16: Virginia, 17: West Virginia

Model Group 4: 1: Alaska, 2: Arizona, [1] 3: California, 4: Colorado, 5: Idaho, 6: Hawaii, 7: Montana, 8: Nevada, 9: New Mexico, 10: Oregon, 11: Utah, 12: Washington, 13: Wyoming

State/Region [b3]

Model Group 1: 1: New York, 2: Pennsylvania, 3: other [1]

Model Group 2: 1: Illinois, 2: Michigan, 3: Ohio, 4: other [1]

Model Group 3: 1: Florida, 2: Texas, 3: other [1]

Model Group 4: 1: California, 2: other [1]

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percent owner occupied.

[3] The States or district assigned to a particular model is based on Census regions.

[b] Binary variable.

[c] Counting variable. A count of all persons in the household.

C.2 Glossary of Terms Used in the Description of the Variables in the Final Model

Factor effect. Represents the effects of levels considered for one-factor, two-factor, and higher order factors.

Reference/reference set. Factor effects composed of reference levels are not explicitly listed in the set of model variables. However, these effects manifest themselves either separately or in combination with other factors depending on the presence of other factors in the model.

All levels present. All effects and all levels of the factor under consideration are in the model.

Coll. *(levels)*. Collapse these factor effects together. Factor effects that have been collapsed with others manifest themselves jointly in the model.

Drop all levels. All factor effects are completely removed from the model for all levels and any combinations involving this factor.

Drop *level(s)*. Collapse these factor effects into the reference set. The factor effects comprising the dropped levels are manifested jointly with either some of or all of the factor effects in the reference set.

Drop *level(s)*; sing. During the modeling process the factor effects listed were removed from the model due to singularity.

Drop *level(s); zero cnts.* During the modeling process the factor effects listed were removed from the model due to zero sample.

Hier. One or more of the effects in a higher order interaction was collapsed or dropped in an interaction at a lower order, either eliminating or combining factors of higher order interactions with that effect.

Do the same for *(effects)*. Repeat the previous step for all effect levels listed.

Drop or Collapse using*. The asterisk is used as a wildcard character to indicate all levels of the factor for that effect.

*Note: The above are given as a list of general terms. Certain other specific terms are sometimes used within a particular section.

C.3 How to Interpret Collapsing and Dropping of Factor Effects

To help visualize what effects are directly controlled for in our model, one can construct the table that reflects the collapsing scheme employed. The following is a complex example from the 2003 person-level modeling.

1. Locate the Factor effect - Model 9 Person Nonresponse Adjustment:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Drop (3,4,2); sing. Collapse (1,4,2) & (1,4,3); conv. Drop (3,*,*); conv. Coll. (4,1,2) & (4,1,3); conv. Repeat for all levels of age in State (4).

2. Determine the initial range of possible levels for the variables by referring to the variable definitions. See Exhibits C.1 and H.1 for QDU- and Pair-level variable definitions. In addition, the columns 'Levels,' 'Proposed,' and 'Final' will provide counts of all factor effects, all explicitly proposed factors, and all explicitly controlled factors, but these are not necessary for construction of the cross-classification table. The following example is based upon person-level variables, but the process is the same.

- State (for the model group in question, in this case, Model Group 9)
Model Group 9: 1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington, 5: California[1,2,3]

- Age
1: 12 to 17,[2,3] 2: 18 to 25, 3: 26 to 34, 4: 35 to 49, 5: 50+ [1]

- Race (3 levels)
1: white,[1,2,3] 2: black, 3: other

Note that superscript number indicates the reference level of the variable for a particular stage of modeling. In our case, the model stage is 'Person Nonresponse Adjustment.'

3. Construct the cross-classification table.

For example, Race (4 Levels) is defined this way:

Race (4 Levels	white	black	Asian	American Indian/Alaska Native

Indicates the reference-level set.

This is the cross-classification table for State × Race (4 Levels):

State*Race (4 Levels)	white	black	Asian	American Indian/Alaska Native
AK				
HI				
OR				
WA				
CA				

Indicates the reference-level set.

The cross-classification table of interest (State × Age × Race [3 Levels]) is as follows:

State × Age × Race (3 Levels)	white	black	other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Indicates the reference-level set.

The number of respondents in that class at this stage of modeling would appear within each cell of the table. Construction of the other cross-classification tables follows the same logic and is only necessary to the point of providing understanding of the final table.

4. Use the information under the 'Comments' column definition to determine the combination of factors controlled.

Hier. This note means the factor effect was collapsed at a lower order. Because this note is present, examine the information on lower-order factor effects that are the components of the interaction term, State × Race (3 levels) × Age; that is, look at the one-factor and two-factor effects for State, Race (4 levels) and Age, and their accompanying information:

One-Factor Effects	**Comments**
State	All levels present.
Race (4 Levels)	All levels present.
Age	All levels present.
Two-Factor Effects	**Comments**
State × Age	All levels present.
State × Race (4 Levels)	Collapse (1,3) & (1,4). Do the same for all other States except (2). Collapse (2,2), (2,3), & (2,4).
Age × Race (3 Levels)	All levels present.

The reason for the note is the State × Race (4 Levels) interaction. It indicates a need to maintain the collapsing scheme when setting up any three-factor crosses involving State × Race. Following these directions, the resulting two-factor table is:

<table>
<tr><th>State × Race (4 Levels)</th><th>white</th><th>black</th><th>Asian</th><th>American Indian/Alaska Native</th></tr>
<tr><td>AK</td><td rowspan="4"></td><td></td><td colspan="2"></td></tr>
<tr><td>HI</td><td colspan="3"></td></tr>
<tr><td>OR</td><td></td><td colspan="2"></td></tr>
<tr><td>WA</td><td></td><td colspan="2"></td></tr>
<tr><td>CA</td><td colspan="4"></td></tr>
</table>

Indicates the reference-level set.

Returning to our instructions, we see that several other factor crosses have been affected by modeling:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Drop (3,4,2); sing. Collapse (1,4,2) & (1,4,3); conv. Drop (3,*,*); conv. Coll. (4,1,2) & (4,1,3); conv. Repeat for all levels of age in State (4).

Construct the complete table, then begin combining blocks as directed. The unshaded cells represent the factors directly controlled for by the model. The shaded cells represent the composite reference set, whose values may be obtained by utilizing the marginal sums, although when changes to the initially proposed set occur, it can make certain reference cell counts indistinguishable.

After following the directions, the cross-classification table should appear as follows:

State × Age × Race (3 Levels)	white	black	other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Indicates the reference-level set.

Exhibit C.2 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights

Variables	Binary	Counting	Level	Proposed
One-Factor Effects		**76**	**76**	
Intercept	Y		1	1
Population density	Y		4	3
Group quarter	Y		3	2
Race of Householder	Y		6	5
Rent/housing value	Y		5	4
Segment % Black	Y		3	2
Segment % Hispanic	Y		3	2
Segment % Owner-Occupied	Y		3	2
Household type	Y		7	6
State	Y	Y	Model Specific	
Quarter	Y	Y	4	3
Age Group		Y	5	4
Race		Y	5	4
Hispanicity		Y	2	1
Gender		Y	2	1
Household size		Y	1	1
Two-Factor Effects				
Age x Race (3 levels)		Y	5 x 3	8
Age x Hispanicity		Y	5 x 2	4
Age x Gender		Y	5 x 2	4
Race (3 levels) x Hispanicity		Y	3 x 2	2
Race (3 levels) x Gender		Y	3 x 2	2
Hispanicity x Gender		Y	2 x 2	1
State x Age		Y	Model Specific	
State x Race (5 levels)		Y	Model Specific	
State x Gender		Y	Model Specific	
State x Hispanicity		Y	Model Specific	
% Black x % Owner	Y		3 x 3	4
% Black x Rent/housing		Y	3 x 5	8
% Hispanicity x % Owner		Y	3 x 3	4
% Hispanicity x Rent/housing		Y	3 x 5	8
% Owner x Rent/housing	Y		3 x 5	8
Three-Factor Effects				
Race (3 levels) x Age x Gender		Y	8	8
State/Region x Age x Gender		Y		
State/Region x Age x Hispanicity		Y		
State/Region x Age x Race (3 levels)		Y		
State/Region x Hispanicity x Gender		Y		
State/Region x Race (3 levels) x Hispanicity		Y		
State/Region x Race (3 levels) x Gender		Y		

Appendix C1: Model Group 1: Northeast

(Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont)

Table C1a 2003 QDU Weight GEM Modeling Summary (Model Group 1: Northeast)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	1.48%	1.38%	0.34%	2.6567	243	(0.66, 2.85)	(0.68, 2.85)
	1.49%	1.94%	0.44%	2.7088	242	(0.21, 2.85)	(0.21, 2.57)
						(0.32, 1.85)	(0.33, 1.77)
res.qdu.nr	1.49%	2.06%	0.46%	2.7386	243	(1.00, 2.50)	(1.00, 2.50)
	1.48%	2.95%	0.48%	3.1843	239	(1.00, 4.60)	(1.00, 4.55)
						(1.20, 1.31)	(1.20, 1.23)
res.qdu.ps	1.48%	2.95%	0.48%	3.1843	243	(0.80, 2.29)	(0.91, 2.27)
	1.51%	2.61%	0.25%	3.1825	239	(0.72, 2.29)	(0.73, 1.42)
						(0.99, 1.02)	(0.99, 1.02)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C1b 2003 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	23	1.00	23	0.21	23	0.52	36	0.69	36
1%	104	1.00	112	0.61	114	1.00	125	0.93	125
5%	142	1.00	175	0.76	167	1.02	185	0.98	184
10%	197	1.00	251	0.81	240	1.04	266	0.99	267
25%	369	1.00	590	0.90	575	1.08	615	0.99	614
Median	721	1.08	959	0.98	950	1.15	1,047	1.00	1,046
75%	1021	3.28	2,053	1.08	2,048	1.26	2,196	1.00	2,215
90%	1390	6.01	4,633	1.20	4,641	1.40	5,609	1.01	5,595
95%	1705	8.11	6,744	1.30	6,719	1.52	8,753	1.02	8,757
99%	2332	12.66	11,659	1.59	11,783	1.82	16,437	1.09	16,446
Maximum	7269	18.15	25,867	3.17	28,011	4.55	35,161	1.64	37,550
n	11,639	-	11,639	-	11,639	-	9,732	-	9,732
Mean	781	2.43	1832	1.00	1,822	1.20	2,179	1.00	2,179
Max/Mean	9.30	-	14.12	-	15.38	-	16.14	-	17.24

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

Model Group 1 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The Northeast model group maintained all originally proposed covariates except for one level of the segment percent Hispanic by segment rent/housing variable, which was removed due to zero counts.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

All main effects were maintained in full for the nonresponse adjustment. Among two-factor effects, the segment level variable percent Hispanicity by percent renting/house was dropped due to zero sample, and the races Native American and Asian in Vermont and New Hampshire were collapsed due to convergence problem. Among three-factor effects, the races non-Hispanic Black and Other in Pennsylvania were collapsed due to convergence problems.

Respondent Questionnaire Dwelling Unit–Level Poststratification

All main effects and three-way interactions were kept in full for the poststratification adjustment. Among non-state two-factor effects, the segment level variable percent Hispanicity by percent renting/housing was dropped due to zero sample. Among state two-factor effects, the races Native American and Multiple Race were collapsed in Maine and Rhoda Island due to convergence problems, and the races Native American and Asian were collapsed in New Hampshire due to zero sample.

Exhibit C1.1 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**132**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	32	All levels present.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	7	Drop (1,2); zero cnts.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**242**	

Exhibit C1.2 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	4	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**130**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	30	Coll (4,3) & (4,4), (9,3) & (9,4); conv.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	7	Drop (1, 2); zero cnts.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**49**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	3	Coll. (7,2) & (7,3); conv.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**239**	

Exhibit C1.3 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	8	8	All levels present.
State (binary)	9	8	8	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**129**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	9 x 5	32	32	All levels present.
State x Race	9 x 5	32	29	Coll. (2,3) & (2,5). Repeat for State (8); conv. Coll. (4,3) & (4,4); zero cnts.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	7	Drop (1,2); zero cnts.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**243**	**239**	

Appendix C2: Model Group 2: Midwest

(Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin)

Table C2a 2003 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	2.16%	1.34%	0.19%	2.5366	300	(0.40, 2.80)	(0.40, 2.60)
	1.61%	1.55%	0.28%	2.4979	300	(0.25, 4.25)	(0.25, 4.25)
						(0.99, 2.28)	(0.99, 2.28)
res.qdu.nr	1.60%	1.50%	0.29%	2.5590	300	(1.00, 2.79)	(1.00, 2.78)
	1.26%	1.12%	0.22%	2.7896	293	(1.00, 4.50)	(1.00, 4.50)
						(1.00, 5.00)	(1.00, 4.77)
res.qdu.ps	1.26%	1.12%	0.22%	2.7896	300	(0.96, 2.21)	(0.98, 2.21)
	1.25%	1.14%	0.14%	2.7885	297	(0.79, 2.28)	(0.80, 1.41)
						(0.99, 1.04)	(0.99, 1.04)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.
[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[3] Number of proposed covariates on top line, and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C2b 2003 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	40	1.00	41	0.22	14	0.62	14	0.68	12
1%	139	1.00	148	0.61	138	1.00	148	0.93	148
5%	181	1.00	236	0.81	235	1.02	247	0.99	247
10%	330	1.00	433	0.86	417	1.04	439	0.99	438
25%	541	1.00	589	0.93	582	1.08	640	1.00	640
Median	637	1.05	796	0.99	798	1.13	890	1.00	892
75%	812	3.02	1,818	1.07	1,811	1.22	1,918	1.00	1,918
90%	1247	5.84	4,253	1.16	4,268	1.35	5,168	1.01	5,160
95%	1442	7.63	5,648	1.25	5,789	1.45	7,245	1.01	7,235
99%	1858	11.95	9,741	1.66	10,227	1.71	13,379	1.05	13,445
Maximum	5018	13.16	24,728	4.49	20,513	4.77	2,8336	1.67	27,440
n	15,542	-	15,542	-	15,542	-	13,288	-	13,288
Mean	719	2.30	1,634	1.01	1,637	1.17	1,914	1.00	1,914
Max/Mean	6.98	-	15.13	-	12.53	-	14.80	-	14.34

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

Model Group 2 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The Midwest model group maintained all originally proposed covariates.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

Insufficient sample for State-level race categories led to the creation of a combined Native American and Multiple Race within Indiana, Iowa, Minnesota, Nebraska, North Dakota, South Dakota, and Kansas. All other proposed factor effects were retained in full.

Respondent Questionnaire Dwelling Unit–Level Poststratification

For the final adjustment for the Midwest model group, all factors except State by race were kept at proposed levels. State by race was altered by combining Native American and Multiple Race within North Dakota, South Dakota, and Minnesota.

Exhibit C2.1 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**163**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	44	All levels present.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**300**	

Exhibit C2.2 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**156**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	37	Coll. (2,3) & (2,5); conv. Repeat for States (3), (4), (6), (8), (9) & (11); conv.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**293**	

Exhibit C2.3 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	9	11	11	All levels present.
State (binary)	9	11	11	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**160**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	12 x 5	44	44	All levels present.
State x Race	12 x 5	44	41	Coll (6,3) & (6,5); conv. Repeat for States (9) & (11).
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	3	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Hispanicity	4 x 5 x 2	12	12	All levels present.
State/Region x Age x Race (3 levels)	4 x 5 x 3	24	24	All levels present.
State/Region x Hispanicity x Gender	4 x 2 x 2	3	3	All levels present.
State/Region x Race (3 levels) x Hispanicity	4 x 3 x 2	6	6	All levels present.
State/Region x Race (3 levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**297**	

Appendix C3: Model Group 3: South

Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia

Table C3a 2003 QDU Weight GEM Modeling Summary (Model Group 3: South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	1.13%	1.73%	0.31%	2.42126	339	(0.54, 2.05)	(0.55, 2.05)
	0.92%	1.55%	0.29%	2.43923	337	(0.30, 3.93)	(0.30, 3.93)
						(0.95, 2.41)	(0.95, 2.41)
res.qdu.nr	0.90%	1.74%	0.33%	2.46865	339	(1.00, 2.93)	(1.00, 2.91)
	0.94%	1.60%	0.27%	2.71447	336	(1.00, 4.07)	(1.00, 3.87)
						(1.00, 2.27)	(1.00, 2.19)
res.qdu.ps	0.94%	1.60%	0.27%	2.71447	339	(0.80, 1.95)	(0.95, 1.95)
	0.91%	1.19%	0.18%	2.71576	337	(0.80, 1.95)	(0.80, 1.84)
						(0.99, 1.25)	(0.99, 1.25)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

[3] Number of proposed covariates on top line, and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C3b 2003 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	20	1.00	29	0.30	17	0.54	17	0.70	16
1%	66	1.00	72	0.64	69	1.00	74	0.94	73
5%	131	1.00	197	0.80	196	1.02	214	0.99	215
10%	279	1.00	430	0.85	413	1.04	446	0.99	447
25%	665	1.00	815	0.92	801	1.08	882	1.00	881
Median	965	1.09	1,282	0.99	1,290	1.14	1,418	1.00	1,420
75%	1322	3.18	2,536	1.07	2,575	1.24	2,738	1.00	2,746
90%	1775	5.93	5,964	1.15	5,776	1.35	6,982	1.01	6,992
95%	2098	7.32	8,259	1.22	8,282	1.44	10,311	1.02	10,315
99%	2661	11.57	13,153	1.50	13,285	1.69	17,257	1.06	17,181
Maximum	6840	16.18	53,992	3.92	39,200	3.87	37,300	1.84	37,900
n	17,194	-	17,194	-	17,194	-	14,676	-	14,676
Mean	1025	2.36	2,324	1.00	2,323	1.18	2,721	1.00	2721
Max/Mean	6.67	-	23.23	-	16.88	-	13.71	-	13.93

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

Model Group 3 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The only changes from the proposed set of initial covariates involved the State by race interaction. The races Native American and Asian were collapsed in Kentucky and Mississippi.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

The only changes from the proposed set of initial covariates involved the State by race interaction. The races Native American and Two or More races was collapsed in Louisiana. The races Native American, Asian, and Two or More races were collapsed in Kentucky.

Respondent Questionnaire Dwelling Unit–Level Poststratification

The only changes from the proposed set of initial covariates involved the State by race interaction. The races Native American and Asian were collapsed in Kentucky and Mississippi.

Exhibit C3.1 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**211**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	62	Coll (7,3) & (7,4); zero cnts. Coll (10,3) & (10,4); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**337**	

Exhibit C3.2 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**210**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	61	Coll. (8,3) & (8,5); conv. Coll. (7,3), (7,4), & (7,5); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**336**	

Exhibit C3.3 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	17	16	16	All levels present.
State (binary)	17	16	16	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**211**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	17 x 5	64	64	All levels present.
State x Race	17 x 5	64	62	Coll (7,3) & (7,4); zero cnts. Coll (10,3) & (10,4); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Hispanicity	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Race (3 levels)	3 x 5 x 3	16	16	All levels present.
State/Region x Hispanicity x Gender	3 x 2 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Hispanicity	3 x 3 x 2	4	4	All levels present.
State/Region x Race (3 levels) x Gender	3 x 3 x 2	4	4	All levels present.
Total		**339**	**337**	

Appendix C4: Model Group 4: West

(Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming)

Table C4a 2003 QDU Weight GEM Modeling Summary (Model Group 4: West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Outwinsor			Nominal	Realized
sel.qdu.ps	1.87%	2.41%	0.50%	3.1142	270	(0.60, 2.75)	(0.60, 2.75)
	1.18%	1.82%	0.51%	3.1788	268	(0.45, 5.00)	(0.45, 4.99)
						(0.95, 1.54)	(0.95, 1.54)
res.qdu.nr	1.41%	3.31%	0.71%	3.22407	270	(1.00, 2.95)	(1.00, 2.95)
	1.12%	2.87%	0.57%	3.71081	268	(1.00, 2.95)	(1.00, 2.83)
						(1.20, 1.27)	(1.20, 1.20)
res.qdu.ps	1.12%	2.87%	0.57%	3.71081	270	(0.90, 2.10)	(0.93, 2.10)
	1.25%	2.81%	0.44%	3.71556	267	(0.30, 2.10)	(0.32, 1.36)
						(0.99, 1.23)	(0.99, 1.23)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.
[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[3] Number of proposed covariates on top line, and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C4b 2003 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)

	SDU wt	QDU design weight		sel.qdu.ps		res.qdu.nr		res.qdu.ps	
	1-9	duwght10	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13
Minimum	21	1.00	21	0.33	25	0.54	28	0.32	29
1%	86	1.00	90	0.66	87	1.00	91	0.92	91
5%	117	1.00	138	0.78	138	1.01	153	0.98	152
10%	142	1.00	192	0.83	190	1.03	210	0.99	210
25%	272	1.00	391	0.91	399	1.07	441	0.99	439
Median	659	1.07	1,059	0.99	1,070	1.13	1,158	1.00	1,161
75%	1551	2.92	2,059	1.09	2,086	1.24	2,319	1.01	2,327
90%	2012	5.26	5,186	1.20	5,102	1.39	5,524	1.01	5,514
95%	2235	7.17	7,690	1.31	7,739	1.49	9,757	1.02	9,725
99%	2842	11.53	14,683	1.58	15,038	1.73	21,085	1.08	21,172
Maximum	8669	14.83	40,905	4.99	39,335	2.83	54,528	1.60	54,550
n	11,809	-	11,809	-	11,809	-	10,057	-	10,057
Mean	927	2.19	1,983	1.01	2,003	1.18	2,352	1.00	2,352
Max/Mean	9.35	-	20.63	-	19.64	-	23.19	-	23.20

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7B.

Model Group 4 Overview

Selected Questionnaire Dwelling Unit–Level Poststratification

The West model group maintained all originally proposed covariates except for two levels of the segment percent black by segment rent/housing variable, which were collapsed due to zero counts.

Respondent Questionnaire Dwelling Unit–Level Nonresponse

The West model group maintained all originally proposed covariates except for two levels of the segment percent black by segment rent/housing variable, which were collapsed due to zero counts.

Respondent Questionnaire Dwelling Unit–Level Poststratification

The West model group maintained all originally proposed covariates except for two levels of the segment percent black by segment rent/housing variable, which were collapsed due to zero counts; and one level for the State by race interaction where the races black and Two or More Races was collapsed in Wyoming.

Exhibit C4.1 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**171**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	48	All levels present.
State x Gender	13 x 2	12	12	All levels present.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	6	Coll (1,1) & (2,1). Coll (1,2) & (2,2); zero cnts.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**268**	

Exhibit C4.2 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**171**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	48	All levels present.
State x Gender	13 x 2	12	12	All levels present.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	6	Coll (1,1) & (2,1). Coll (1,2) & (2,2); zero cnts.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**268**	

Exhibit C4.3 Covariates for 2003 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group quarter	3	2	2	All levels present.
Race of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black	3	2	2	All levels present.
% Hispanic	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (count)	13	12	12	All levels present.
State (binary)	13	12	12	All levels present.
Quarter (count)	4	3	3	All levels present.
Quarter (binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**170**	
Age x Race (3 levels)	5 x 3	8	8	All levels present.
Age x Hispanicity	5 x 2	4	4	All levels present.
Age x Gender	5 x 2	4	4	All levels present.
Race (3 levels) x Hispanicity	3 x 2	2	2	All levels present.
Race (3 levels) x Gender	3 x 2	2	2	All levels present.
Hispanicity x Gender	2 x 2	1	1	All levels present.
State x Age	13 x 5	48	48	All levels present.
State x Race	13 x 5	48	47	Coll (13,2) & (13,5), conv.
State x Gender	13 x 2	12	12	All levels present.
State x Hispanicity	13 x 2	12	12	All levels present.
% Black x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black x Rent/Housing	3 x 5	8	6	Coll (1,1) & (2,1). Coll (1,2) & (2,2); zero cnts.
% Hispanicity x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity x Rent/housing	3 x 5	8	8	All levels present.
% Owner-Occupied x Rent/housing	3 x 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 levels) x Age x Gender	3 x 5 x 2	8	8	All levels present.
State/Region x Age x Gender	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Hispanicity	2 x 5 x 2	4	4	All levels present.
State/Region x Age x Race (3 levels)	2 x 5 x 3	8	8	All levels present.
State/Region x Hispanicity x Gender	2 x 2 x 2	1	1	All levels present.
State/Region x Race (3 levels) x Hispanicity	2 x 3 x 2	2	2	All levels present.
State/Region x Race (3 levels) x Gender	2 x 3 x 2	2	2	All levels present.
Total		**270**	**267**	

Appendix D: Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Response Rates

Table D.1 2003 NSDUH QDU-Level Response Rates

Domain	Selected QDU	Respondent QDU	Interview Response Rate [1]
Total	56,184	47,753	79.51%
Census Region			
Northeast	11,639	9,732	77.22%
South	17,194	14,676	80.47%
Midwest	15,542	13,288	80.36%
West	11,809	10,057	79.00%
Quarter			
Quarter 1	13,787	11,686	78.79%
Quarter 2	14,285	12,163	79.72%
Quarter 3	14,659	12,480	80.05%
Quarter 4	13,453	11,424	79.45%
Household Type			
12-17, 18-25, 26+	4,845	4,409	90.68%
12-17, 18-25	136	128	94.63%
12-17, 26+	16,730	15,051	89.82%
18-25, 26+	10,892	9,188	83.84%
12-17	98	83	85.80%
18-25	7,108	6,330	88.50%
26+	16,375	12,564	75.31%
Race of Householder			
Hispanic white	6,118	5,281	83.65%
Hispanic black	122	111	90.01%
Hispanic other	331	289	85.72%
Non-Hispanic white	39,477	33,376	78.74%
Non-Hispanic black	6,679	5,816	81.26%
Non-Hispanic other	3,457	2,880	77.38%
% Hispanic in Segment			
50-100%	2,979	2,564	83.64%
10-50%	9,202	7,747	77.90%
<10%	44,003	37,442	79.59%
% Black in Segment			
50-100%	4,399	3,836	81.30%
10-50%	7,811	6,704	81.18%
<10%	43,974	37,213	79.01%
% Owner-Occupied DUs in Segment			
50-100%	42,093	35,644	79.20%
10-50%	10,691	9,145	79.36%
<10%	3,400	2,964	84.16%
Combined Median Rent/Housing Value			
1st Quintile	10,767	9,248	81.29%
2nd Quintile	11,077	9,501	80.54%
3rd Quintile	11,959	10,131	78.37%
4th Quintile	10,959	9,235	78.70%
5th Quintile	11,422	9,638	79.35%
Population Density			
Large MSA	20,124	16,804	77.88%
Medium-Small MSA	20,834	17,864	80.76%
Non-MSA, Urban	6,928	6,032	81.46%
Non-MSA, Rural	8,298	7,053	80.34%
Group Quarters			
Group	1,034	983	94.29%
Nongroup	55,150	46,770	79.38%
Household Size			
One	7,437	6,039	76.64%
Two	20,885	17,196	77.62%
Three	15,733	13,631	83.84%
Four or more	12,129	10,887	88.20%

[1] The weight used for calculating the response rate includes SDU-level and QDU-level design weights, SDU nonresponse and poststratification adjustments, and selected QDU poststratification adjustment. This weight is the product of YR03WT1*...*YR03WT9*DU03WT10*DU03WT11

Appendix E: Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Proportions of Extreme Values and Outwinsors

(This page intentionally left blank)

Table E.1 2003 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors

Domain	n	Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9) Unweighted	Weighted[2]	Outwinsor[3]	Before sel.qdu.ps[1] (SDUWT*DU03WT10) Unweighted	Weighted[2]	Outwinsor[3]	After sel.qdu.ps[1] (SDUWT*DU03WT10*DU03WT11) Unweighted	Weighted[2]	Outwinsor[3]
Total	56,184	2.22%	3.46%	0.68%	1.64%	1.72%	0.33%	1.30%	1.74%	0.37%
Census Region										
Northeast	11,639	2.04%	4.22%	0.99%	1.48%	1.38%	0.34%	1.49%	1.94%	0.44%
South	17,194	1.61%	2.87%	0.59%	1.13%	1.73%	0.31%	0.97%	1.70%	0.31%
Midwest	15,542	2.79%	3.28%	0.39%	2.16%	1.34%	0.19%	1.61%	1.55%	0.28%
West	11,809	2.54%	3.96%	0.85%	1.87%	2.41%	0.50%	1.18%	1.82%	0.51%
Quarter										
Quarter 1	13,787	2.34%	3.88%	0.80%	1.82%	2.04%	0.39%	1.33%	1.62%	0.38%
Quarter 2	14,285	2.37%	3.59%	0.76%	1.83%	2.35%	0.40%	1.30%	2.25%	0.44%
Quarter 3	14,659	1.75%	2.73%	0.53%	1.26%	1.05%	0.23%	1.02%	1.30%	0.31%
Quarter 4	13,453	2.45%	3.61%	0.62%	1.69%	1.41%	0.29%	1.56%	1.77%	0.35%
Household Type										
12-17, 18-25, 26+	4,845	1.80%	3.41%	0.88%	1.80%	3.41%	0.88%	1.92%	4.58%	1.40%
12-17, 18-25	136	5.88%	6.56%	1.51%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12-17, 26+	16,730	2.28%	3.58%	0.79%	2.24%	3.55%	0.75%	1.73%	3.71%	0.98%
18-25, 26+	10,892	2.31%	3.36%	0.57%	2.30%	3.43%	0.56%	1.84%	4.24%	1.08%
12-17	98	1.02%	1.12%	0.66%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
18-25	7,108	2.45%	3.51%	0.58%	2.27%	3.18%	0.51%	1.51%	2.19%	0.39%
26+	16,375	2.10%	3.39%	0.61%	0.31%	0.86%	0.16%	0.25%	0.74%	0.07%
Race of Householder										
Hispanic white	6,118	2.42%	2.99%	0.59%	1.88%	1.79%	0.35%	1.29%	2.04%	0.53%
Hispanic black	122	54.10%	74.66%	28.18%	42.62%	41.28%	14.84%	36.89%	29.74%	10.66%
Hispanic other	331	21.15%	27.30%	5.85%	13.60%	11.11%	3.28%	7.25%	19.56%	7.99%
Non-Hispanic white	39,477	0.91%	1.32%	0.19%	0.68%	0.92%	0.13%	0.51%	0.78%	0.08%
Non-Hispanic black	6,679	3.01%	5.36%	1.04%	2.46%	2.73%	0.58%	2.95%	3.59%	0.79%
Non-Hispanic other	3,457	11.72%	16.44%	2.91%	8.04%	6.97%	1.26%	5.29%	7.03%	1.87%
% Hispanic in Segment										
50-100%	2,979	2.15%	4.05%	0.99%	1.34%	2.83%	0.49%	1.41%	3.12%	0.59%
10-50%	9,202	2.45%	4.38%	1.12%	1.91%	2.26%	0.53%	1.75%	2.12%	0.62%
<10%	44,003	2.18%	3.14%	0.52%	1.61%	1.50%	0.26%	1.20%	1.54%	0.29%

Table E.1 2003 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors (continued)

		Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)			Before sel.qdu.ps[1] (SDUWT*DU03WT10)			After sel.qdu.ps[1] (SDUWT*DU03WT10*DU03WT11)		
Domain	***n***	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**
% Black in Segment										
50-100%	4,399	3.05%	5.47%	1.12%	2.57%	2.22%	0.41%	2.61%	3.59%	0.75%
10-50%	7,811	3.06%	5.29%	1.10%	2.18%	2.31%	0.50%	2.19%	2.80%	0.67%
<10%	43,974	1.99%	2.85%	0.54%	1.46%	1.56%	0.29%	1.01%	1.35%	0.28%
% Owner-Occupied DUs in Segment										
50-100%	42,093	1.82%	2.68%	0.48%	1.33%	1.42%	0.24%	1.04%	1.49%	0.29%
10-50%	10,691	3.24%	5.46%	1.20%	2.40%	2.56%	0.58%	1.69%	1.98%	0.50%
<10%	3,400	4.00%	6.15%	1.33%	3.12%	3.01%	0.64%	3.32%	4.27%	1.01%
Combined Median Rent/Housing Value										
1st Quintile	10,767	2.27%	2.55%	0.45%	1.64%	0.90%	0.16%	1.23%	1.33%	0.30%
2nd Quintile	11,077	2.44%	2.94%	0.47%	1.96%	1.53%	0.19%	1.55%	1.56%	0.26%
3rd Quintile	11,959	2.68%	4.49%	0.80%	1.89%	2.50%	0.41%	1.55%	2.12%	0.47%
4th Quintile	10,959	2.04%	3.17%	0.68%	1.41%	1.28%	0.29%	1.27%	1.56%	0.43%
5th Quintile	11,422	1.66%	3.75%	0.87%	1.30%	2.11%	0.51%	0.89%	1.97%	0.35%
Population Density										
Large MSA	20,124	2.53%	4.20%	0.82%	1.85%	2.07%	0.39%	1.51%	2.03%	0.46%
Medium-Small MSA	20,834	1.97%	3.06%	0.64%	1.51%	1.75%	0.35%	1.31%	1.79%	0.33%
Non-MSA, Urban	6,928	2.53%	3.50%	0.61%	1.95%	1.31%	0.20%	1.21%	1.23%	0.37%
Non-MSA, Rural	8,298	1.83%	1.71%	0.27%	1.22%	0.69%	0.14%	0.86%	0.95%	0.17%
Group Quarters										
Group	1,034	3.38%	6.06%	1.45%	2.42%	3.87%	0.93%	2.51%	3.56%	0.68%
Nongroup	55,150	2.20%	3.42%	0.67%	1.63%	1.70%	0.32%	1.28%	1.72%	0.37%
Household Size										
One	7,437	1.59%	2.81%	0.61%	0.89%	0.84%	0.16%	0.48%	0.69%	0.08%
Two	20,885	2.23%	3.32%	0.60%	1.32%	1.47%	0.27%	0.99%	1.21%	0.19%
Three	15,733	2.24%	3.47%	0.78%	2.08%	2.88%	0.59%	1.56%	3.03%	0.79%
Four or more	12,129	2.57%	3.99%	0.72%	2.10%	3.26%	0.58%	2.00%	4.65%	1.24%

[1] Sel = selected, QDU = questionnaire dwelling unit, PS = poststratification adjustment.
[2] Weighted extreme value proportion: 100 * $\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: 100 * $\sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table E.2 2003 NSDUH Respondent DU-Level Proportions of Extreme Values and Outwinsors

Domain	n	Before res.qdu.nr[1] (SDUWT*DU03WT10*DU03WT11)			After res.qdu.nr[1] (SDUWT*DU03WT10*...*DU03WT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DU03WT10*...*DU03WT13)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	47,753	1.32%	2.08%	0.43%	1.18%	2.02%	0.36%	1.20%	1.80%	0.24%
Census Region										
Northeast	9,732	1.49%	2.06%	0.46%	1.48%	2.95%	0.48%	1.51%	2.61%	0.25%
South	14,676	0.90%	1.74%	0.33%	0.94%	1.60%	0.27%	0.91%	1.19%	0.18%
Midwest	13,288	1.60%	1.50%	0.29%	1.26%	1.12%	0.22%	1.25%	1.14%	0.14%
West	10,057	1.41%	3.31%	0.71%	1.12%	2.87%	0.57%	1.25%	2.81%	0.44%
Quarter										
Quarter 1	11,686	1.41%	2.24%	0.47%	1.33%	2.43%	0.44%	1.31%	2.20%	0.27%
Quarter 2	12,163	1.29%	2.43%	0.49%	1.18%	1.98%	0.42%	1.19%	1.83%	0.28%
Quarter 3	12,480	1.03%	1.60%	0.35%	0.98%	1.82%	0.30%	1.06%	1.55%	0.20%
Quarter 4	11,424	1.58%	2.04%	0.40%	1.25%	1.86%	0.29%	1.25%	1.62%	0.21%
Household Type										
12-17, 18-25, 26+	4,409	2.02%	4.49%	1.30%	1.50%	4.12%	1.00%	1.56%	4.00%	0.72%
12-17, 18-25	128	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
12-17, 26+	15,051	1.65%	3.57%	0.96%	1.55%	3.88%	0.95%	1.59%	4.05%	0.69%
18-25, 26+	9,188	1.68%	4.10%	1.13%	1.53%	4.17%	1.06%	1.76%	4.29%	0.76%
12-17	83	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
18-25	6,330	1.42%	2.24%	0.42%	1.15%	1.88%	0.23%	1.06%	1.79%	0.19%
26+	12,564	0.41%	1.18%	0.11%	0.40%	1.20%	0.11%	0.28%	0.83%	0.05%
Race of Householder										
Hispanic white	5,281	1.27%	2.08%	0.59%	1.10%	2.42%	0.58%	1.10%	2.42%	0.47%
Hispanic black	111	40.54%	33.33%	11.61%	30.63%	24.09%	6.41%	32.43%	25.31%	5.55%
Hispanic other	289	7.27%	22.55%	9.25%	6.92%	16.76%	4.77%	5.54%	22.66%	5.43%
Non-Hispanic white	33,376	0.55%	1.11%	0.12%	0.43%	0.80%	0.07%	0.38%	0.46%	0.03%
Non-Hispanic black	5,816	2.87%	3.79%	0.71%	2.61%	5.20%	0.93%	2.84%	5.16%	0.59%
Non-Hispanic other	2,880	5.17%	7.43%	2.16%	5.45%	8.63%	1.99%	5.97%	8.65%	1.30%
% Hispanic in Segment										
50-100%	2,564	1.40%	3.23%	0.66%	1.01%	2.03%	0.42%	1.25%	2.48%	0.48%
10-50%	7,747	1.79%	2.78%	0.72%	2.21%	4.19%	0.72%	2.40%	3.65%	0.48%
<10%	37,442	1.22%	1.82%	0.34%	0.98%	1.48%	0.27%	0.95%	1.29%	0.16%

Table E.2 2003 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU03WT10*DU03WT11)			After res.qdu.nr[1] (SDUWT*DU03WT10*...*DU03WT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DU03WT10*...*DU03WT13)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
% Black in Segment										
50-100%	3,836	2.69%	3.43%	0.72%	2.61%	3.94%	0.77%	2.84%	3.87%	0.41%
10-50%	6,704	2.13%	3.13%	0.75%	2.00%	3.60%	0.64%	2.06%	3.28%	0.47%
<10%	37,213	1.04%	1.73%	0.33%	0.88%	1.53%	0.27%	0.88%	1.31%	0.18%
% Owner-Occupied DUs in Segment										
50-100%	35,644	1.08%	1.88%	0.36%	0.94%	1.71%	0.31%	0.95%	1.42%	0.21%
10-50%	9,145	1.71%	2.24%	0.49%	1.85%	3.11%	0.52%	1.91%	3.03%	0.33%
<10%	2,964	3.10%	4.06%	1.05%	1.99%	2.68%	0.52%	2.06%	2.82%	0.39%
Combined Median Rent/Housing Value										
1st Quintile	9,248	1.19%	1.31%	0.35%	1.15%	1.42%	0.27%	1.08%	1.19%	0.17%
2nd Quintile	9,501	1.51%	1.58%	0.26%	1.00%	1.35%	0.22%	1.12%	1.35%	0.17%
3rd Quintile	10,131	1.66%	3.16%	0.62%	1.51%	2.55%	0.48%	1.50%	2.24%	0.28%
4th Quintile	9,235	1.36%	2.07%	0.50%	1.20%	2.37%	0.45%	1.21%	2.20%	0.31%
5th Quintile	9,638	0.88%	2.00%	0.36%	1.02%	2.13%	0.33%	1.07%	1.75%	0.24%
Population Density										
Large MSA	16,804	1.64%	2.79%	0.55%	1.52%	2.75%	0.43%	1.59%	2.48%	0.28%
Medium-Small MSA	17,864	1.28%	1.76%	0.36%	1.11%	1.56%	0.35%	1.16%	1.43%	0.23%
Non-MSA, Urban	6,032	1.16%	1.47%	0.44%	0.99%	1.31%	0.30%	0.83%	1.14%	0.25%
Non-MSA, Rural	7,053	0.81%	0.96%	0.19%	0.71%	1.25%	0.18%	0.68%	0.90%	0.12%
Group Quarters										
Group	983	2.75%	5.08%	0.84%	0.92%	2.40%	0.28%	1.02%	2.55%	0.13%
Nongroup	46,770	1.29%	2.05%	0.42%	1.18%	2.02%	0.36%	1.20%	1.79%	0.24%
Household Size										
One	6,039	0.55%	0.62%	0.11%	0.65%	0.84%	0.11%	0.53%	0.54%	0.04%
Two	17,196	1.02%	1.93%	0.26%	0.98%	1.90%	0.22%	0.92%	1.48%	0.14%
Three	13,631	1.50%	2.93%	0.75%	1.21%	2.61%	0.62%	1.34%	2.87%	0.44%
Four or more	10,887	2.02%	4.55%	1.27%	1.75%	4.59%	1.19%	1.84%	4.69%	0.88%

[1] Res = Respondent, QDU = questionnaire dwelling unit, NR = nonresponse adjustment, PS = poststratification adjustment.
[2] Weighted extreme value proportion: 100 * $\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: 100 * $\sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Appendix F: Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Slippage Rates

(This page intentionally left blank)

Table F.1 2003 NSDUH Questionnaire Dwelling Unit–Level Slippage Rates

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control from SDU Weights (C)	(I-C)/C%	(F-C)/C %
Total	47,753	110,226,196	110,226,196	110,226,196	0.00	-0.00
Census Region						
Northeast	9,732	21,201,845	21,201,845	21,201,845	0.00	-0.00
South	14,676	39,938,143	39,938,143	39,938,143	0.00	-0.00
Midwest	13,288	25,435,639	25,435,639	25,435,639	0.00	-0.00
West	10,057	23,650,569	23,650,569	23,650,569	0.00	0.00
Quarter						
Quarter 1	11,686	27,472,853	27,472,853	27,472,853	0.00	0.00
Quarter 2	12,163	27,562,277	27,562,277	27,562,277	0.00	-0.00
Quarter 3	12,480	27,426,051	27,426,051	27,426,051	0.00	0.00
Quarter 4	11,424	27,765,015	27,765,015	27,765,015	0.00	-0.00
Household Type						
12-17, 18-25, 26+	4,409	4,528,825	4,528,825	4,528,825	0.00	0.00
12-17, 18-25	128	95,425	95,425	95,425	0.00	0.00
12-17, 26+	15,051	14,144,627	14,144,627	14,144,627	0.00	-0.00
18-25, 26+	9,188	11,765,652	11,765,652	11,765,652	0.00	-0.00
12-17	83	53,237	53,237	53,237	0.00	0.00
18-25	6,330	6,460,749	6,460,749	6,460,749	0.00	-0.00
26+	12,564	73,177,682	73,177,682	73,177,682	0.00	0.00
Race of Householder						
Hispanic white	5,281	10,505,884	10,505,884	10,505,884	-0.00	0.00
Hispanic black	111	436,640	436,640	436,640	0.00	0.00
Hispanic other	289	416,498	416,498	416,498	0.00	-0.00
Non-Hispanic white	33,376	80,297,856	80,297,856	80,297,856	0.00	0.00
Non-Hispanic black	5,816	12,861,693	12,861,693	12,861,693	0.00	0.00
Non-Hispanic other	2,880	5,707,624	5,707,624	5,707,624	0.00	0.00
% Hispanic in Segment						
50-100%	2,564	6,239,230	6,239,230	6,239,230	0.00	0.00
10-50%	7,747	20,731,038	20,731,038	20,731,038	0.00	-0.00
<10%	37,442	83,255,928	83,255,928	83,255,928	0.00	-0.00
% Black in Segment						
50-100%	3,836	8,175,461	8,175,461	8,175,461	0.00	0.00
10-50%	6,704	16,667,811	16,667,811	16,667,811	0.00	-0.00
<10%	37,213	85,382,924	85,382,924	85,382,924	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	35,644	83,447,924	83,447,924	83,447,924	0.00	-0.00
10-50%	9,145	20,742,884	20,742,884	20,742,884	0.00	-0.00
<10%	2,964	6,035,388	6,035,388	6,035,388	0.00	0.00
Combined Median Rent/Housing Value						
1st Quintile	9,248	17,081,668	17,081,668	17,081,668	0.00	-0.00
2nd Quintile	9,501	19,727,389	19,727,389	19,727,389	0.00	0.00
3rd Quintile	10,131	23,984,764	23,984,764	23,984,764	0.00	0.00
4th Quintile	9,235	24,777,506	24,777,506	24,777,506	0.00	0.00
5th Quintile	9,638	24,654,869	24,654,869	24,654,869	0.00	0.00

Table F.1 2003 NSDUH Questionnaire Dwelling Unit-Level Slippage Rates (continued)

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control from SDU Weights (C)	(I-C)/C%	(F-C)/C %
Population Density						
Large MSA	16,804	48,726,674	48,726,674	48,726,674	0.00	0.00
Medium-Small MSA	17,864	37,060,022	37,060,022	37,060,022	0.00	-0.00
Non-MSA, Urban	6,032	10,844,006	10,844,006	10,844,006	0.00	0.00
Non-MSA, Rural	7,053	13,595,493	13,595,493	13,595,493	0.00	-0.00
Group Quarters						
Group	983	928,045	928,045	928,045	0.00	0.00
Nongroup	46,770	109,298,152	109,298,151	109,298,151	0.00	-0.00
Household Size						
One	6,039	29,465,712	29,456,318	29,446,192	0.07	0.03
Two	17,196	51,266,648	51,257,893	51,289,036	-0.04	-0.06
Three	13,631	17,348,682	17,371,745	17,394,891	-0.27	-0.13
Four or more	10,887	12,145,154	12,140,240	12,096,077	0.41	0.37

[1] YR03WT1*...*YR03WT9*DU03WT10*...*DU03WT12 (before QDU poststratification).
[2] YR03WT1*...*YR03WT9*DU03WT10*...*DU03WT13 (after QDU poststratification).

Appendix G: Evaluation of Calibration Weights: Questionnaire Dwelling Unit–Level Weight Summary Statistics

(This page intentionally left blank)

Table G.1 2003 NSDUH Selected QDU-Level Weight Summary Statistics

Domain	*n*	Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)						Before sel.qdu.ps[1] (SDUWT*DU03WT10)						After sel.qdu.ps[1] (SDUWT*DU03WT10*DU03WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	56,184	20	503	742	1,159	8,669	1.44	21	603	1,040	2,101	53,992	2.68	14	596	1,038	2,122	39,335	2.71
Census Region																			
Northeast	11,639	23	369	721	1,021	7,269	1.42	23	590	959	2,053	25,867	2.66	23	575	950	2,048	28,011	2.71
South	17,194	20	665	965	1,322	6,840	1.32	29	815	1,282	2,536	53,992	2.42	17	801	1,290	2,575	39,200	2.44
Midwest	15,542	40	541	637	812	5,018	1.26	41	589	796	1,818	24,728	2.54	14	582	798	1,811	20,513	2.50
West	11,809	21	272	659	1,551	8,669	1.68	21	391	1,059	2,059	40,905	3.11	25	399	1,070	2,086	39,335	3.18
Quarter																			
Quarter 1	13,787	24	512	766	1,178	7,739	1.46	24	628	1,024	2,185	40,905	2.77	30	607	1,025	2,154	39,335	2.73
Quarter 2	14,285	21	501	722	1,132	8,669	1.46	21	592	1,036	2,123	53,992	2.70	14	580	1,026	2,121	39,200	2.73
Quarter 3	14,659	23	498	706	1,117	6,690	1.42	23	589	1,019	2,001	26,237	2.59	23	587	1,011	2,055	26,346	2.63
Quarter 4	13,453	20	502	771	1,218	6,121	1.43	24	618	1,089	2,095	27,178	2.65	17	623	1,101	2,157	28,286	2.71
Household Type																			
12-17, 18-25, 26+	4,845	22	539	790	1,252	8,669	1.46	22	539	790	1,252	8,669	1.46	20	525	785	1,247	8,479	1.52
12-17, 18-25	136	24	294	612	1,000	2,791	1.52	24	294	612	1,000	2,791	1.52	17	275	593	982	3,029	1.61
12-17, 26+	16,730	21	480	696	1,109	7,362	1.46	21	486	705	1,117	7,407	1.47	14	465	711	1,120	8,812	1.51
18-25, 26+	10,892	23	552	803	1,242	6,012	1.42	24	603	907	1,436	7,073	1.43	23	586	909	1,439	15,698	1.51
12-17	98	43	135	336	806	2,334	1.94	43	135	347	814	2,394	1.95	51	143	286	956	2,241	1.93
18-25	7,108	20	443	730	1,110	7,269	1.42	29	477	811	1,240	7,380	1.43	24	464	791	1,238	6,831	1.45
26+	16,375	30	513	730	1,156	5,519	1.43	113	2,020	3,658	6,037	53,992	1.61	87	2,002	3,624	5,924	39,335	1.63
Race of Householder																			
Hispanic white	6,118	25	538	892	1,433	6,690	1.36	25	630	1,153	1,886	24,728	2.22	17	621	1,154	1,923	23,530	2.31
Hispanic black	122	43	1,031	1,647	2,279	8,669	1.62	43	1,220	2,056	4,155	28,174	2.51	60	1,340	2,190	4,591	37,695	2.48
Hispanic other	331	23	180	414	865	7,739	2.60	24	230	544	1,358	13,318	3.46	23	196	478	1,303	15,297	3.58
Non-Hispanic white	39,477	21	503	715	1,096	5,018	1.41	21	604	1,039	2,207	53,992	2.70	14	600	1,036	2,242	39,335	2.72
Non-Hispanic black	6,679	24	599	843	1,259	7,362	1.39	24	692	1,060	1,998	28,312	2.59	36	668	1,052	1,994	26,000	2.66
Non-Hispanic other	3,457	20	259	580	1,228	6,840	1.86	30	327	841	1,920	26,902	3.04	23	329	845	1,919	30,898	2.95
% Hispanic in Segment																			
50-100%	2,979	72	692	1,137	1,590	7,739	1.27	73	823	1,430	2,097	19,992	2.20	36	801	1,406	2,139	23,530	2.24
10-50%	9,202	21	570	995	1,486	8,669	1.39	21	744	1,342	2,394	28,174	2.49	17	738	1,347	2,445	37,695	2.52
<10%	44,003	20	472	693	1,056	7,269	1.44	22	582	964	2,030	53,992	2.76	14	573	963	2,047	39,335	2.78

Table G.1 2003 NSDUH Selected QDU-Level Weight Summary Statistics (continued)

		Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)						Before sel.qdu.ps[1] (SDUWT*DU03WT10)						After sel.qdu.ps[1] (SDUWT*DU03WT10*DU03WT11)					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
% Black in Segment																			
50-100%	4,399	20	564	795	1,169	6,174	1.43	32	643	972	1,874	28,174	2.77	23	621	977	1,897	37,695	2.86
10-50%	7,811	35	606	893	1,331	6,840	1.36	40	723	1,186	2,183	27,178	2.48	17	731	1,236	2,258	30,898	2.52
<10%	43,974	21	460	709	1,124	8,669	1.46	21	583	1,023	2,105	53,992	2.71	14	575	1,011	2,114	39,335	2.73
%Owner-Occupied DUs in Segment																			
50-100%	42,093	21	502	723	1,125	8,669	1.44	21	599	1,030	2,148	53,992	2.69	23	593	1,023	2,166	39,200	2.72
10-50%	10,691	24	497	797	1,238	7,739	1.47	24	610	1,091	2,057	40,905	2.67	14	605	1,080	2,059	39,335	2.68
<10%	3,400	20	523	853	1,244	7,269	1.42	24	625	1,043	1,811	26,631	2.56	43	614	1,099	1,902	24,623	2.56
Combined Median Rent/Housing Value																			
1st Quintile	10,767	21	367	606	822	6,840	1.44	21	531	800	1,748	25,399	2.75	23	517	800	1,747	28,286	2.81
2nd Quintile	11,077	22	339	656	1,021	5,263	1.55	22	526	896	1,924	26,902	2.89	14	506	905	1,883	30,898	2.89
3rd Quintile	11,959	33	533	769	1,119	7,269	1.41	33	627	1,054	2,158	40,905	2.71	43	622	1,083	2,161	39,335	2.69
4th Quintile	10,959	37	634	964	1,352	8,669	1.33	37	761	1,252	2,369	22,683	2.38	17	740	1,248	2,476	27,899	2.46
5th Quintile	11,422	20	548	855	1,314	7,362	1.41	29	678	1,215	2,280	53,992	2.63	24	670	1,208	2,345	39,200	2.63
Population Density																			
Large MSA	20,124	20	666	968	1,404	8,669	1.30	29	804	1,360	2,693	40,905	2.35	24	795	1,371	2,732	39,335	2.37
Medium-Small MSA	20,834	25	414	676	1,043	7,739	1.45	25	553	943	1,861	53,992	2.80	17	547	948	1,882	39,200	2.82
Non-MSA, Urban	6,928	21	261	579	921	4,986	1.58	21	378	793	1,623	20,588	3.04	14	377	778	1,586	21,753	3.07
Non-MSA, Rural	8,298	30	289	596	917	5,018	1.54	33	439	813	1,712	25,894	2.99	23	436	798	1,689	27,785	3.03
Group Quarters																			
Group	1,034	24	258	584	1,035	7,269	1.65	24	294	665	1,169	18,549	2.81	24	279	650	1,211	12,134	2.46
Nongroup	55,150	20	508	744	1,161	8,669	1.44	21	608	1,050	2,132	53,992	2.67	14	602	1,048	2,151	39,335	2.69
Household Size																			
One	7,437	37	470	703	1,072	7,269	1.42	64	862	2,208	5,976	53,992	2.17	45	860	2,221	5,838	39,335	2.19
Two	20,885	20	495	729	1,142	7,362	1.42	21	720	1,458	3,395	30,509	2.12	26	707	1,450	3,360	25,418	2.16
Three	15,733	22	515	748	1,169	6,840	1.44	22	545	829	1,340	13,715	1.92	23	530	824	1,343	15,406	2.00
Four or more	12,129	23	519	781	1,251	8,669	1.48	23	525	799	1,301	11,004	1.61	14	508	794	1,307	19,222	1.72

[1] Sel = selected, QDU = questionnaire dwelling unit, PS = poststratification.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table G.2 2003 NSDUH Respondent QDU-Level Weight Summary Statistics

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU03WT10*DU03WT11)						After res.qdu.nr[1] (SDUWT*DU03WT10*...*DU03WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU03WT10*...*DU03WT13)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	47,753	14	582	995	1,953	39,335	2.75	14	650	1,139	2,298	54,528	3.08	12	651	1,142	2,302	54,550	3.08
Census region																			
Northeast	9,732	23	555	910	1,804	28,011	2.74	36	615	1,047	2,196	35,161	3.18	36	614	1,046	2,215	37,550	3.18
South	14,676	17	781	1,254	2,305	30,898	2.47	17	882	1,418	2,738	37,300	2.71	16	881	1,420	2,746	37,900	2.72
Midwest	13,288	14	574	773	1,638	20,513	2.56	14	640	890	1,918	28,336	2.79	12	640	892	1,918	27,440	2.79
West	10,057	26	381	1,007	1,992	39,335	3.22	28	441	1,158	2,319	54,528	3.71	29	439	1,161	2,327	54,550	3.72
Quarter																			
Quarter 1	11,686	30	594	988	1,972	39,335	2.78	30	666	1,131	2,322	54,528	3.16	34	667	1,130	2,329	54,550	3.16
Quarter 2	12,163	14	567	983	1,965	37,695	2.77	14	630	1,122	2,312	37,696	3.07	12	631	1,126	2,310	45,627	3.07
Quarter 3	12,480	23	576	970	1,894	26,346	2.67	38	641	1,110	2,252	35,965	2.98	37	641	1,108	2,247	35,894	2.98
Quarter 4	11,424	17	601	1,050	1,997	28,286	2.75	17	675	1,213	2,320	37,300	3.10	16	677	1,215	2,326	37,550	3.10
Household Type																			
12-17, 18-25, 26+	4,409	20	525	783	1,245	8,479	1.51	20	565	861	1,367	8,513	1.53	19	567	862	1,368	8,017	1.51
12-17, 18-25	128	17	262	593	1,010	3,029	1.63	17	285	651	1,051	3,189	1.59	16	286	649	1,050	3,173	1.59
12-17, 26+	15,051	14	463	711	1,121	8,812	1.51	14	508	789	1,246	10,822	1.52	12	510	789	1,250	8,651	1.50
18-25, 26+	9,188	23	581	901	1,437	15,698	1.52	29	667	1,061	1,706	18,040	1.55	26	666	1,062	1,708	13,535	1.53
12-17	83	51	143	331	956	2,241	1.94	51	155	383	1,119	2,668	1.93	51	153	378	1,112	2,673	1.92
18-25	6,330	24	455	786	1,227	6,831	1.45	24	495	887	1,399	6,303	1.45	24	494	887	1,399	6,307	1.45
26+	12,564	87	1,947	3,550	5,795	39,335	1.64	88	2,459	4,529	7,701	54,528	1.70	84	2,460	4,527	7,704	54,550	1.70
Race of Householder																			
Hispanic white	5,281	17	617	1,139	1,879	23,530	2.33	17	691	1,277	2,121	36,439	2.62	16	688	1,278	2,126	36,348	2.62
Hispanic black	111	60	1,326	2,115	4,335	37,695	2.62	80	1,347	2,266	4,727	37,696	2.63	75	1,312	2,302	4,211	45,627	2.97
Hispanic other	289	23	196	456	1,254	15,297	3.73	29	229	575	1,550	25,889	4.04	26	223	569	1,509	27,523	4.19
Non-Hispanic white	33,376	14	586	988	2,022	39,335	2.77	14	655	1,136	2,416	54,528	3.08	12	656	1,137	2,419	54,550	3.08
Non-Hispanic black	5,816	36	657	1,020	1,887	19,509	2.62	36	726	1,132	2,161	35,161	3.08	36	723	1,132	2,164	37,550	3.08
Non-Hispanic other	2,880	23	311	789	1,779	30,898	3.08	24	364	908	2,191	36,717	3.44	24	365	911	2,183	37,900	3.46
% Hispanic in Segment																			
50-100%	2,564	36	791	1,389	2,082	23,530	2.25	38	917	1,515	2,362	36,439	2.56	37	917	1,513	2,355	36,348	2.57
10-50%	7,747	17	710	1,286	2,218	37,695	2.55	17	804	1,496	2,667	37,696	2.94	16	804	1,498	2,681	45,627	2.94
<10%	37,442	14	560	922	1,866	39,335	2.83	14	624	1,053	2,204	54,528	3.15	12	623	1,053	2,203	54,550	3.15

Table G.2 2003 NSDUH Respondent QDU-Level Weight Summary Statistics (continued)

		Before res.qdu.nr[1] (SDUWT*DU03WT10*DU03WT11)						After res.qdu.nr[1] (SDUWT*DU03WT10*...*DU03WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU03WT10*...*DU03WT13)					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
% Black in Segment																			
50-100%	3,836	23	612	950	1,785	37,695	2.83	38	679	1,082	2,053	37,696	3.17	35	683	1,081	2,065	45,627	3.20
10-50%	6,704	17	714	1,189	2,134	30,898	2.52	17	800	1,343	2,513	37,300	2.80	16	801	1,346	2,528	37,900	2.80
<10%	37,213	14	560	965	1,933	39,335	2.78	14	625	1,109	2,290	54,528	3.13	12	625	1,110	2,291	54,550	3.13
% Owner-Occupied DUs in Segment																			
50-100%	35,644	23	580	980	1,984	37,695	2.76	24	648	1,122	2,350	37,696	3.09	24	650	1,122	2,350	45,627	3.09
10-50%	9,145	14	586	1,027	1,910	39,335	2.71	14	655	1,189	2,244	54,528	3.09	12	655	1,189	2,245	54,550	3.08
<10%	2,964	43	606	1,066	1,833	24,623	2.58	45	659	1,205	2,061	33,511	2.86	45	656	1,202	2,064	33,350	2.87
Combined Median Rent/Housing Value																			
1st Quintile	9,248	23	503	776	1,604	28,286	2.88	32	558	877	1,920	37,300	3.12	32	558	877	1,919	37,401	3.11
2nd Quintile	9,501	14	493	871	1,733	30,898	2.93	14	560	996	1,992	36,717	3.21	12	559	998	1,995	37,900	3.21
3rd Quintile	10,131	43	609	1,028	1,918	39,335	2.74	43	683	1,188	2,268	54,528	3.12	40	684	1,189	2,273	54,550	3.13
4th Quintile	9,235	17	722	1,195	2,202	22,764	2.50	17	815	1,375	2,635	35,161	2.86	16	815	1,376	2,646	37,550	2.86
5th Quintile	9,638	24	649	1,148	2,190	28,011	2.65	24	733	1,319	2,604	34,023	2.96	24	733	1,320	2,614	34,024	2.96
Population Density																			
Large MSA	16,804	24	771	1,304	2,476	39,335	2.41	24	882	1,503	3,029	54,528	2.72	24	883	1,505	3,034	54,550	2.72
Medium-Small MSA	17,864	17	533	913	1,756	30,898	2.85	17	592	1,041	2,034	36,717	3.18	16	592	1,040	2,041	37,900	3.17
Non-MSA, Urban	6,032	14	368	752	1,473	21,753	3.14	14	412	856	1,684	28,226	3.48	12	413	857	1,685	29,742	3.50
Non-MSA, Rural	7,053	23	422	771	1,575	27,785	3.05	32	481	880	1,849	33,340	3.36	32	481	878	1,859	33,256	3.36
Group Quarters																			
Group	983	24	281	645	1,207	12,134	2.50	24	297	655	1,250	16,981	2.77	24	295	650	1,249	16,982	2.76
Nongroup	46,770	14	588	1,004	1,981	39,335	2.73	14	658	1,152	2,340	54,528	3.06	12	659	1,153	2,342	54,550	3.06
Household Size																			
One	6,039	45	815	1,930	5,504	39,335	2.27	45	964	2,339	7,053	54,528	2.38	66	963	2,356	7,057	54,550	2.38
Two	17,196	26	677	1,346	3,037	24,721	2.20	32	775	1,580	3,843	29,926	2.41	30	772	1,581	3,844	29,742	2.41
Three	13,631	23	521	812	1,318	14,710	1.95	28	580	921	1,503	19,884	2.22	26	581	921	1,505	21,420	2.23
Four or more	10,887	14	503	788	1,289	19,222	1.69	14	553	872	1,435	26,447	1.87	12	553	870	1,439	27,261	1.85

[1] Res = respondent, QDU = questionnaire dwelling unit, PS = poststratification adjustment, NR = nonresponse adjustment
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Appendix H: GEM Modeling Summary for the Pair Weights

Appendix H: GEM Modeling Summary for the Pair Weights

Introduction

This appendix summarizes each model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this section provides a model-specific overview of weight calibration, as opposed to a domain-specific one.

For 2003, modeling involved taking two model groups through four adjustment steps: 1) selected pair poststratification, 2) pair nonresponse adjustment, 3) responding pair poststratification, and 4) responding pair extreme value adjustment.

Model-specific summary statistics are shown in Tables H1a, H1b to H2a, and H2b. Included in these tables, for each stage of modeling, are: the number of factor effects included in the final model; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; the weighted, unweighted and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data lead to situations whereby modeling inclusion of all originally proposed levels of covariates in the model is not possible. The text and exhibits in Sections H1 and H2 summarize the decisions made with regard to final covariates included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit H.2; for the list of realized final model covariates, see Exhibits H1.1 to H2.4. For guidelines on interpreting these exhibits, see Appendix C.

H.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 4.2 (included here as Exhibit H.1 for easy reference). There, a complete list is provided of all variables and associated levels used at any stage of modeling. Note that not all factors or levels are present in all stages of modeling, and the initial set of variables is the same across model groups but may change for an adjustment step of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment.

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables

Group Quarter Indicator
1: College Dorm, 2: Other Group Quarter, 3: Nongroup Quarter [1]

Household Size
2: DU with 2 persons, 3: DU with 3 persons, 4: DU with >=4 persons

Pair Age (15 levels)
1: 12-17 & 12-17 [1], 2: 12-17 & 18-25, 3: 12-17 & 26-34, 4: 12-17 & 35-49, 5: 12-17 & 50+,
6: 18-25 & 18-25, 7: 18-25 & 26-34, 8: 18-25 & 35-49, 9: 18-25 & 50+, 10: 26-34 & 26-34,
11: 26-34 & 35-49, 12: 26-34 & 50+, 13: 35-49 & 35-49, 14: 35-49 & 50+, 15: 50+ & 50+

Pair Age (6 levels)
1: 12-17 & 12-17 [1], 2: 12-17 & 18-25, 3: 12-17 & 26+, 4: 18-25 & 18-25, 5: 18-25 & 26+, 6: 26+ & 26+.

Pair Age (3 levels)
1: 12-17 & 12-17 [1], 2: 12-17 & 18+, 3: 18+ & 18+

Pair Gender
1: Male & Female [1], 2: Female & Female, 3: Male & Male

Pair Race (10 levels)
1: white & white [1], 2: white & black, 3: white & Hispanic, 4: white & other, 5: black & black,
6: black & Hispanic, 7: black & other, 8: Hispanic & Hispanic, 9: Hispanic & other, 10: other & other.

Pair Race (5 levels)
1: Mixed race pair, 2: Hispanic pair, 3: black pair, 4: white pair[1], 5: other pair.

Pair Race (4 levels)
1: Mixed race pair or other & other, 2: Hispanic pair, 3: black pair, 4 white pair[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)
1: 50%-100% [1], 2: 10%->50%, 3: 0->10%

Percentage of Segments That Are Black (% black)
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Percentage of Segments That Are Hispanic (% Hispanic)
1: 50%-100%, 2: 10%->50%, 3: 0->10% [1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

Population Density
1: MSA 1,000,000 or more, 2: MSA less than 1,000,000, 3: Non-MSA urban, 4: Non-MSA rural [1]

Quarter
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4 [1]

Race of Householder
1: Hispanic white [1], 2: Hispanic black, 3: Hispanic others, 4: Non-Hispanic white,
5: Non-Hispanic black, 6: Non-Hispanic others,

State / Region

Model Group 1**:** 1: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, Rhode Island, Vermont, 2: Alabama, Arkansas, Delaware, District of Columbia, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Virginia, West Virginia [1], 3: New York, 4: Pennsylvania, 5: Florida, 6: Texas,

Model Group 2: 1: Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota,| Wisconsin, 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming [1], 3: Michigan, 4: Illinois, 5: Ohio, 6: California

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables (continued)

States [3]

Model Group 1: 1: Alabama, 2: Arkansas, 3: Connecticut, 4: Delaware, 5: District of Columbia, 6: Florida, 7: Georgia, 8: Kentucky, 9: Louisiana, 10: Maine, 11: Maryland, [1] 12: Massachusetts, 13: Mississippi, 14: New Hampshire, 15: New Jersey, 16: New York, 17: North Carolina, 18: Oklahoma, 19: Pennsylvania, 20: Rhode Island, 21: South Carolina, 22: Tennessee, 23: Texas, 24: Vermont, 25: Virginia, 26: West Virginia

Model Group 2: 1: Alaska, 2: Arizona, [1] 3: California, 4: Colorado, 5: Idaho, 6: Illinois, 7: Indiana, 8: Iowa, 9: Hawaii, 10: Kansas, 11: Michigan, 12: Minnesota, 13: Missouri, 14: Montana, 15: Nebraska, 16: Nevada, 17: New Mexico, 18: North Dakota, 19: Ohio, 20: Oregon, 21: South Dakota, 22: Utah, 23: Washington, 24: Wisconsin, 25: Wyoming

Pair Relationship Associated with Multiplicity

1: Parent-child (12-14)*
2: Parent-child (12-17)*
3: Parent-child (12-10)*
4: Parent*-child (12-14)
5: Parent*-child (12-17)
6: Parent*-child (12-20)
7: Sibling (12-14)-sibling (15-17)
8: Sibling (12-17)-sibling (18-25)
9: Spouse-spouse
10: Spouse-spouse with kids

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.
[2] Segment-combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percent Owner-Occupied.
[3] The States or district assigned to a particular model is based on combined Census regions.

Exhibit H.2 Covariates for 2003 NSDUH Pair Weights

Variables	Level	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Population Density	3	2
Group Quarter	3	2
Household size	3	2
Pair Age	15	14
Pair Sex	4	2
Pair Race	10	9
Race of Householder	6	5
Rent/housing	5	4
Segment % black	3	2
Segment % Hispanic	3	2
% Owner-Occupied	3	2
Pair Relationship	Model Specific	
Two-Factor Effects		
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20
Pair Race (5 levels) x Pair Sex	5 x 3	8
Pair Sex x Pair Age (6 levels)	3 x 6	10
State/Region x Pair Race (5 levels)	Model Specific	
State/Region x Pair Age (6 levels)	Model Specific	
State/Region x Pair Sex	Model Specific	
Rent/housing x % black	5 x 3	8
Rent/housing x % Hispanic	5 x 3	8
Rent/housing x % Owner-Occupied	5 x 3	8
% Owner-Occupied x % black	3 x 3	4
% Owner-Occupied x % Hispanic	3 x 3	4
Three-Factor Effects		
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12

Appendix H1: Model Group 1: Northeast and South

(Alabama, Arkansas, Connecticut, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maine, Massachusetts, Maryland, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, West Virginia)

Table H1a   2003 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Winsorized			Nominal	Realized
sel.pr.ps	6.37%	24.0%	11.80%	78.943	213	(0.2, 1.20)	(0.20, 1.20)
	1.39%	5.51%	0.98%	9.359	205	(0.2, 5.00)	(0.20, 4.42)
						(0.9, 1.41)	(0.90, 1.41)
res.pr.nr	1.45%	5.86%	1.15%	10.001	213	(1.00, 2.40)	(1.00, 2.40)
	2.01%	10.30%	2.36%	10.321	213	(1.00, 5.00)	(1.00, 5.00)
						n/a	n/a
res.pr.ps	2.09%	10.20%	2.59%	10.321	223	(0.63, 1.30)	(0.63, 1.30)
	1.82%	6.84%	0.47%	10.451	219	(0.25, 2.20)	(0.25, 2.20)
						n/a	n/a
res.pr.ev	1.82%	6.84%	0.47%	10.451	223	(0.50, 1.30)	(0.95, 1.24)
	0.17%	1.23%	0.04%	10.465	219	(0.50, 2.00)	(0.81, 1.39)
						(0.93, 1.81)	(1.81, 1.81)

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7A.
[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[3] Number of proposed covariates on top line, and number finalized after modeling.
[4] Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second for nonextreme, and the third for low extreme values.

Table H1b 2003 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Midwest and West)

	SDU wt	pair selection prob		sel.pr.ps [1]		res.pr.nr [1]		res.pr.ps [1]		res.pr.ev [1]	
	1-9	pairwt10	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14
Minimum	20	1.02	29	0.03	14	0.33	14	0.20	4	0.63	4
1%	65	1.02	154	0.21	106	1.00	107	0.32	77	0.89	76
5%	144	1.15	331	0.36	293	1.00	298	0.53	240	0.94	234
10%	230	1.21	652	0.49	512	1.01	529	0.70	438	0.96	435
25%	574	1.41	1,233	0.73	1,178	1.05	1,217	0.89	1,155	0.98	1,140
Median	886	4.72	3,133	1.03	3,196	1.13	3,491	0.99	3,422	1.00	3,415
75%	1,241	10.08	8,566	1.42	8,686	1.28	9,746	1.09	9,722	1.01	9,728
90%	1,709	18.03	18,176	1.88	19,507	1.60	24,170	1.22	24,072	1.03	24,081
95%	2,055	26.05	26,803	2.22	29,333	1.91	40,514	1.32	40,961	1.05	41,916
99%	2,725	45.68	56,512	2.94	64,491	2.97	103,231	2.19	97,3343	1.09	96,946
Maximum	6,174	4,339.03	7,860,600	4.42	1,049,310	5.00	1,162,890	2.20	1,127,000	1.81	1,138,770
n	12,628	-	12,628	-	12,628	-	9,859	-	9,859	-	9,859
mean	955	8.61	8,713	1.12	8,339	1.25	10,682	0.99	10,682	1.00	10,682
Max/mean	6.46	-	902.2	-	125.8	-	108.9	-	105.5	-	106.6

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7A.

Model Group 1 Overview

Selected Pair–Level Poststratification

All 76 proposed main effects were included in the model.

In addition, all 125 proposed two-factor effects were included in the model.

The three-factor interaction of pair race by pair sex by pair age was simplified by collapsing. Here pair race categories mixed and other pair, Hispanic pair, and black pair were collapsed for all combinations of age and gender. As a result, out of 12 three-factor effects, 4 were kept in the model.

Respondent Pair–Level Nonresponse

All 213 proposed factors were retained in the final model.

Respondent Pair–Level Poststratification

All 86 proposed main effects were included in the model.

Also, all 125 proposed two-factor effects were retained in the model.

The three-factor interaction of pair race by pair sex by pair age was simplified by collapsing. Here the mixed pair race category was collapsed with black pair race category for all 12 proposed variables. As a result, out of 12 three-factor effects, 8 were kept in the model.

Respondent Pair–Level Extreme Value Adjustment

This step used exactly the same variables as in the respondent pair–level poststratification step.

Exhibit H1.1 Covariates for 2003 NSDUH Pair Weights (sel.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**4**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	4	Coll. (1,1,1), (2,1,1), & (3,1,1); conv. Repeat for all levels of pair sex and pair age.
Total		**213**	**205**	

Exhibit H1.2 Covariates for 2003 NSDUH Pair Weights (res.pr.nr) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	12	All levels present.
Total		**213**	**213**	

Exhibit H1.3 Covariates for 2003 NSDUH Pair Weights (res.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**86**	**86**	
Intercept	1	1	1	All levels present.
State	26	24	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**8**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	8	Coll. (1,1,2) & (3,1,2); conv. Repeat for all levels of pair sex, pair age.
Total		**223**	**219**	

Exhibit H1.4 Covariates for 2003 NSDUH Pair Weights (res.pr.ev) Model Group 1: Northeast and South

This step used the same variables as the respondent pair–level poststratification step in Exhibit H1.3.

Appendix H2: Model Group 2: Midwest and West

(Alaska, Arizona, California, Colorado, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oregon, South Dakota, Utah, Washington, Wisconsin, Wyoming)

Table H2a   2003 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	Unweighted	Weighted	Winsorized			Nominal	Realized
sel.pr.ps	6.74%	22.80%	10.32%	78.567	212	(0.22, 1.2)	(0.22, 1.20)
	1.54%	9.67%	2.08%	18.820	199	(0.20, 3.76)	(0.20, 3.74)
						(0.90, 1.74)	(0.90, 1.74)
res.pr.nr	1.60%	11.00%	2.40%	20.994	212	(1.00, 2.30)	(1.00, 2.30)
	2.23%	16.58%	5.46%	20.296	212	(1.00, 5.00)	(1.00, 5.00)
						(1.42, 1.43)	(1.42, 1.42)
res.pr.ps	2.31%	19.10%	6.95%	20.297	222	(0.55, 1.10)	(0.55, 1.10)
	1.86%	7.49%	0.95%	12.230	210	(0.30, 2.28)	(0.30, 2.28)
						n/a	n/a
res.pr.ev	1.86%	7.49%	0.95%	12.230	222	(0.5, 1.10)	(0.80, 1.10)
	0.07%	0.53%	0.02%	12.224	210	(0.5, 2.0)	(0.74, 1.39)
						n/a	n/a

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7A.
[2] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[3] Number of proposed covariates on top line, and number finalized after modeling.
[4] Nominal bounds are used in defining maximum / minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second for nonextreme, and the third for low extreme values.

Table H2b 2003 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

	SDU wt	pair selection		sel.pr.ps [1]		res.pr.nr [1]		res.pr.ps [1]		res.pr.ev [1]	
	1-9	pairwt10	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14
Minimum	22	1.02	29	0.04	23	0.49	23	0.18	8	0.29	8
1%	90	1.02	159	0.21	107	0.95	104	0.30	40	0.80	38
5%	144	1.12	305	0.33	228	1.00	222	0.34	116	0.91	112
10%	202	1.21	507	0.43	392	1.00	387	0.38	209	0.93	207
25%	447	1.43	932	0.66	901	1.00	904	0.59	667	0.97	656
Median	648	4.40	2,561	0.96	2,334	1.02	2,378	0.89	2,150	0.97	2,152
75%	1,098	9.79	7,112	1.37	6,514	1.17	6,863	1.22	7,071	1.03	6,971
90%	1,753	15.57	15,481	1.84	16,088	1.58	19,398	1.59	20,038	1.06	19,863
95%	2013	25.45	21,486	2.15	26,963	2.13	32,390	1.82	35,962	1.10	36,408
99%	2,610	45.19	51,646	2.81	63,895	4.18	105,859	2.15	111,286	1.17	111,701
Maximum	8,669	4,881.43	7,272,550	4.05	1,721,630	5.00	2,254,660	2.28	1,085,330	1.39	1,047,510
n	12,819	-	12,819	-	12,819	-	10,172	-	10,172	-	10,172
mean	828	8.50	7,521	1.07	7,284	1.22	9,180	0.95	9,180	1.00	9,180
Max/mean	10.47	-	966.90	-	236.4	-	245.6	-	118.2	-	114.1

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7A.

Model Group 2 Overview

Selected Pair–Level Poststratification

All 75 proposed main effects were included in the model.

In addition, 124 of 125 two-factor effects were retained in the model. One State/region by pair race factor was collapsed.

However, none of the 12 three-factor effects were kept in the model due to convergence problems.

Respondent Pair–Level Nonresponse

All 212 proposed factor effects were retained at proposed levels.

Respondent Pair–Level Poststratification

All 85 proposed main effects were included in the model.

All 125 proposed two-factor effects were kept in the model.

None of the 12 three-factor effects were kept in the model due to the problems with convergence.

Respondent Pair–Level Extreme Value Adjustment

This step used the exactly same variables as in the respondent pair–level poststratification step.

Exhibit H2.1 Covariates for 2003 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**124**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	19	Coll. (7,1) & (7,2); conv.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**212**	**199**	

Exhibit H2.2 Covariates for 2003 NSDUH Pair Weights (res.pr.nr) Model Group 2: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	26	25	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	12	All levels present.
Total		**212**	**212**	

Exhibit H2.3 Covariates for 2003 NSDUH Pair Weights (res.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**85**	**85**	
Intercept	1	1	1	All levels present.
State	26	24	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Sex	3	2	2	All levels present.
Pair Race	10	9	9	All levels present.
Race of Householder	6	5	5	All levels present.
Rent/housing	5	4	4	All levels present.
Segment % black	3	2	2	All levels present.
Segment % Hispanic	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race (5 levels) x Pair Age (6 levels)	5 x 6	20	20	All levels present.
Pair Race (5 levels) x Pair Sex	5 x 3	8	8	All levels present.
Pair Sex x Pair Age (6 levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race (5 levels)	6 x 5	20	20	All levels present.
State/Region x Pair Age (6 levels)	6 x 6	25	25	All levels present.
State/Region x Pair Sex	6 x 3	10	10	All levels present.
Rent/housing x % black	5 x 3	8	8	All levels present.
Rent/housing x % Hispanic	5 x 3	8	8	All levels present.
Rent/housing x % Owner-Occupied	5 x 3	8	8	All levels present.
% Owner-Occupied x % black	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race (4 levels) x Pair Sex x Pair Age (3 levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**222**	**210**	

Exhibit H2.4 Covariates for 2003 NSDUH Pair Weights (res.pr.ev) Model Group 2: Northeast and South

This step used the same variables as the respondent pair–level poststratification step in Exhibit H2.3.

Appendix I: Evaluation of Calibration Weights: Pair–Level Response Rates

(This page intentionally left blank)

Table I.1 2003 NSDUH Pair-Level Response Rates

Domain	Selected Pair Size	Respondent Pair Size	Interview Response Rate [1]
Total	25,447	20,031	70.64%
Pair Age Group			
12-17, 12-17	4,302	3,870	90.79%
12-17, 18-25	3,206	2,675	83.92%
12-17, 26-34	927	766	81.70%
12-17, 35-49	3,892	3,168	81.18%
12-17, 50+	584	444	76.57%
18-25, 18-25	5,522	4,362	77.36%
18-25, 26-34	1,085	804	69.40%
18-25, 35-49	1,533	1,087	72.40%
18-25, 50+	695	449	65.53%
26-34, 26-34	856	600	69.46%
26-34, 35-49	493	345	61.46%
26-34, 50+	195	106	58.18%
35-49, 35-49	892	594	71.16%
35-49, 50+	392	222	61.36%
50+, 50+	873	539	56.00%
Pair Race			
Hispanic	3,271	2,518	69.25%
black	2,669	2,196	78.88%
white	16,306	12,855	71.16%
other	1,387	1,016	49.98%
white & black	202	165	66.69%
white & Hispanic	753	602	80.83%
white & other	593	468	71.94%
black & Hispanic	80	62	68.91%
black & other	89	70	78.82%
Hispanic & other	97	79	63.26%
Pair Gender			
Male, Male	5,568	4,333	69.61%
Female, Female	5,347	4,460	75.61%
Male, Female	14,532	11,238	69.58%
Household Size			
Two	7,392	5,540	66.12%
Three	6,947	5,414	67.11%
Four or more	11,108	9,077	75.00%

Table I.1 2003 NSDUH Person Pair-Level Response Rates (continued)

Domain	Selected Pairs	Respondent Pairs	Interview Response Rate [1]
Census Region			
Northeast	5,097	3,923	67.30%
South	7,531	5,936	72.87%
Midwest	7,123	5,705	74.72%
West	5,696	4,467	66.61%
Quarter			
Quarter 1	6,237	4,919	73.15%
Quarter 2	6,345	5,039	70.28%
Quarter 3	6,736	5,245	69.96%
Quarter 4	6,129	4,828	69.21%
% Hispanic in Segment			
50-100%	1,605	1,242	68.68%
10-50%	4,222	3,242	68.42%
<10%	19,620	15,547	71.55%
% Black in Segment			
50-100%	1,914	1,569	78.64%
10-50%	3,460	2,709	72.98%
<10%	20,073	15,753	69.36%
% Owner-Occupied DUs in Segment			
50-100%	19,452	15,266	70.57%
10-50%	4,573	3,625	70.38%
<10%	1,422	1,140	76.38%
Combined Median Rent/Housing Value			
1st Quintile	4,754	3,873	75.68%
2nd Quintile	4,896	3,880	75.36%
3rd Quintile	5,542	4,364	70.95%
4th Quintile	5,047	3,869	67.18%
5th Quintile	5,208	4,045	67.36%
Population Density			
Large MSA	9,187	6,996	67.38%
Medium-Small MSA	9,409	7,544	73.71%
Non-MSA, Urban	3,136	2,561	75.50%
Non-MSA, Rural	3,715	2,930	71.95%
Group Quarters			
Group	545	483	86.03%
Nongroup	24,902	19,548	70.58%

[1] The weight used for calculating the response rate includes SDU-level and pair-level design weights, SDU nonresponse and poststratification adjustments, and selected pair poststratification adjustment. This weight is the product of YR03WT1*..*YR03WT9*PR03WT10*PR03WT11.

Appendix J: Evaluation of Calibration Weights: Pair–Level Proportions of Extreme Values and Outwinsors

(This page intentionally left blank)

Table J.1 2003 NSDUH Selected Pair–Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)			Before sel.pr.ps[1] (SDUWT*PR03WT10)			After sel.pr.ps[1] (SDUWT*PR03WT10*PR03WT11)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	25,447	1.64%	3.24%	0.69%	6.68%	31.93%	22.15%	1.61%	15.31%	9.40%
Pair Age Group										
12-17, 12-17	4,302	1.12%	2.33%	0.47%	6.60%	23.25%	9.25%	0.63%	2.63%	0.50%
12-17, 18-25	3,206	1.31%	2.97%	0.59%	12.13%	35.15%	14.92%	0.84%	3.63%	0.50%
12-17, 26-34	927	2.27%	4.63%	1.28%	1.51%	6.87%	2.51%	0.76%	3.84%	0.49%
12-17, 35-49	3,892	1.41%	2.87%	0.73%	2.62%	10.10%	2.69%	0.64%	1.89%	0.25%
12-17, 50+	584	0.86%	2.43%	0.71%	1.54%	6.74%	2.09%	0.68%	0.82%	0.04%
18-25, 18-25	5,522	1.68%	3.25%	0.63%	10.01%	35.29%	15.73%	1.99%	7.65%	1.19%
18-25, 26-34	1,085	3.04%	4.63%	0.58%	2.03%	7.43%	2.52%	1.20%	5.49%	0.49%
18-25, 35-49	1,533	2.41%	4.11%	0.74%	6.07%	20.16%	6.25%	3.26%	7.75%	1.16%
18-25, 50+	695	1.58%	4.72%	1.82%	3.31%	13.71%	5.83%	1.29%	2.96%	0.49%
26-34, 26-34	856	3.04%	5.72%	1.29%	3.74%	26.91%	15.64%	3.04%	26.54%	12.89%
26-34, 35-49	493	2.23%	3.95%	0.86%	9.33%	53.53%	38.29%	6.69%	42.90%	25.49%
26-34, 50+	195	2.05%	4.42%	0.40%	4.10%	42.63%	35.94%	2.56%	22.33%	15.65%
35-49, 35-49	892	1.57%	2.32%	0.39%	6.61%	50.71%	38.03%	3.92%	44.51%	32.88%
35-49, 50+	392	1.53%	1.99%	0.11%	4.85%	41.47%	32.42%	1.79%	22.59%	17.82%
50+, 50+	873	1.37%	2.37%	0.46%	5.38%	54.61%	49.00%	3.55%	19.93%	14.26%
Pair Race										
Hispanic	3,271	2.90%	5.29%	1.52%	6.91%	36.64%	24.26%	2.32%	19.34%	10.82%
black	2,669	1.87%	3.65%	0.60%	8.65%	32.89%	20.19%	2.29%	17.56%	9.18%
white	16,306	0.36%	0.76%	0.10%	6.07%	29.91%	21.96%	1.01%	13.27%	9.28%
other	1,387	8.94%	16.00%	2.45%	9.59%	29.76%	11.21%	3.17%	18.28%	6.88%
white & black	202	2.97%	3.43%	0.56%	6.44%	20.80%	4.48%	6.93%	14.14%	2.74%
white & Hispanic	753	1.99%	3.20%	0.87%	6.37%	32.99%	23.99%	1.86%	20.75%	13.22%
white & other	593	6.58%	11.05%	2.02%	6.07%	36.54%	28.97%	2.87%	13.31%	8.85%
black & Hispanic	80	22.50%	49.76%	21.46%	16.25%	58.42%	35.20%	11.25%	24.01%	7.97%
black & other	89	7.87%	23.94%	5.43%	10.11%	81.75%	71.66%	6.74%	51.56%	21.95%
Hispanic & other	97	6.19%	13.71%	1.51%	2.06%	1.37%	0.04%	3.09%	2.66%	0.43%
Pair Gender										
Male, Male	5,568	1.76%	3.76%	0.78%	9.03%	30.51%	15.94%	2.10%	9.67%	3.07%
Female, Female	5,347	1.87%	3.60%	0.57%	7.26%	16.43%	6.04%	1.98%	9.61%	2.19%
Male, Female	14,532	1.51%	2.91%	0.69%	5.57%	35.76%	27.39%	1.28%	18.42%	13.11%
Household Size										
Two	7,392	1.23%	1.79%	0.32%	0.46%	1.17%	0.28%	0.41%	2.11%	0.52%
Three	6,947	1.74%	3.61%	0.91%	2.85%	29.03%	22.00%	1.96%	14.74%	8.43%
Four or more	11,108	1.85%	3.90%	0.77%	13.22%	47.94%	32.60%	2.19%	22.74%	14.72%

Table J.1 2003 NSDUH Selected Pair–Level Proportions of Extreme Values and Outwinsors (continued)

		Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)			Before sel.pr.ps[1] (SDUWT*PR03WT10)			After sel.pr.ps[1] (SDUWT*PR03WT10*PR03WT11)		
Domain	***n***	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**
Census Region										
Northeast	5,097	1.45%	4.08%	1.04%	6.65%	39.33%	31.30%	1.96%	18.09%	10.37%
South	7,531	1.38%	2.66%	0.51%	6.48%	25.61%	14.60%	1.26%	12.35%	7.57%
Midwest	7,123	1.99%	2.93%	0.48%	7.50%	30.15%	19.97%	1.71%	16.46%	10.09%
West	5,696	1.72%	3.77%	0.87%	5.95%	35.26%	25.90%	1.62%	16.27%	10.57%
Quarter										
Quarter 1	6,237	1.57%	3.10%	0.73%	6.85%	31.02%	19.46%	1.46%	17.76%	12.20%
Quarter 2	6,345	1.89%	3.71%	0.78%	6.26%	36.19%	27.51%	1.47%	15.24%	8.38%
Quarter 3	6,736	1.41%	3.05%	0.68%	6.38%	23.97%	14.44%	1.59%	13.05%	7.43%
Quarter 4	6,129	1.71%	3.11%	0.55%	7.28%	35.31%	25.76%	1.93%	15.28%	9.68%
% Hispanic in Segment										
50-100%	1,605	2.18%	4.34%	1.18%	6.67%	27.55%	15.58%	2.24%	15.48%	7.78%
10-50%	4,222	2.27%	4.81%	1.22%	6.63%	29.50%	17.42%	2.37%	17.85%	11.06%
<10%	19,620	1.46%	2.66%	0.47%	6.69%	33.08%	24.16%	1.39%	14.53%	9.11%
% Black in Segment										
50-100%	1,914	2.25%	5.10%	1.27%	9.98%	46.42%	31.87%	2.87%	25.70%	14.34%
10-50%	3,460	2.54%	4.89%	1.12%	7.98%	29.15%	16.60%	2.17%	16.91%	9.42%
<10%	20,073	1.43%	2.71%	0.54%	6.14%	30.78%	22.01%	1.39%	13.93%	8.89%
% Owner-Occupied DUs in Segment										
50-100%	19,452	1.33%	2.62%	0.50%	6.49%	33.21%	24.06%	1.50%	15.15%	9.83%
10-50%	4,573	2.45%	4.96%	1.22%	7.13%	25.20%	12.88%	2.36%	16.42%	7.76%
<10%	1,422	3.31%	5.54%	1.33%	7.81%	31.33%	17.88%	0.63%	12.28%	4.92%
Combined Median Rent/Housing Value										
1st Quintile	4,754	1.66%	2.64%	0.50%	7.61%	26.10%	14.17%	1.39%	15.62%	10.36%
2nd Quintile	4,896	1.14%	2.00%	0.39%	7.07%	32.77%	23.63%	1.21%	13.13%	8.76%
3rd Quintile	5,542	1.89%	3.55%	0.69%	7.29%	31.76%	20.06%	1.26%	19.97%	13.26%
4th Quintile	5,047	1.86%	3.84%	0.97%	6.42%	35.13%	26.17%	1.68%	15.09%	8.83%
5th Quintile	5,208	1.61%	3.64%	0.72%	5.07%	31.20%	23.09%	2.48%	12.33%	6.06%
Population Density										
Large MSA	9,187	2.18%	4.17%	0.92%	7.04%	38.50%	28.63%	2.35%	17.24%	9.86%
Medium-Small MSA	9,409	1.24%	2.37%	0.51%	6.29%	25.46%	15.68%	1.26%	14.47%	10.38%
Non-MSA, Urban	3,136	1.69%	3.25%	0.52%	6.79%	20.15%	10.92%	1.21%	10.06%	4.73%
Non-MSA, Rural	3,715	1.29%	2.05%	0.40%	6.68%	27.09%	17.33%	0.97%	13.57%	8.28%
Group Quarters										
Group	545	1.47%	2.43%	0.58%	4.59%	13.53%	4.38%	1.83%	7.67%	1.79%
Nongroup	24,902	1.65%	3.26%	0.69%	6.73%	31.99%	22.21%	1.60%	15.34%	9.43%

[1] This step used demographic variables from screener data for all selected person pairs; Sel = selected, PR = pair, PS = poststratification adjustment.

[2] Weighted extreme value proportion: 100 * $\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.

[3] Outwinsor weight proportion: 100 * $\sum_k(w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table J.2 2003 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors

Domain	n	Before res.pr.nr[1] (SDUWT*PR03WT10*PR03WT11)			After res.pr.nr[1] (SDUWT*PR03WT10*...*PR03WT12)		
		Unweighted	Weighted[2]	Outwinsor[3]	Unweighted	Weighted[2]	Outwinsor[3]
Total	20,031	1.69%	16.88%	9.96%	2.25%	21.51%	10.25%
Pair Age Group							
12-17, 12-17	3,870	0.72%	2.97%	0.55%	0.13%	0.83%	0.10%
12-17, 18-25	2,675	0.90%	3.66%	0.57%	0.86%	4.29%	1.26%
12-17, 26-34	766	0.39%	3.71%	0.33%	0.52%	2.09%	1.06%
12-17, 35-49	3,168	0.79%	2.15%	0.34%	0.88%	3.65%	0.82%
12-17, 50+	444	0.68%	0.71%	0.15%	1.35%	3.64%	0.97%
18-25, 18-25	4,362	1.99%	7.49%	1.19%	2.66%	13.58%	3.58%
18-25, 26-34	804	1.24%	3.27%	0.34%	2.99%	10.10%	2.34%
18-25, 35-49	1,087	3.77%	10.18%	1.47%	4.97%	13.18%	2.40%
18-25, 50+	449	1.34%	2.32%	0.27%	2.90%	11.73%	2.71%
26-34, 26-34	600	3.00%	28.65%	15.56%	5.83%	42.30%	17.20%
26-34, 35-49	345	5.80%	33.63%	17.53%	11.30%	54.05%	22.73%
26-34, 50+	106	7.55%	45.39%	28.91%	5.66%	31.22%	16.46%
35-49, 35-49	594	5.56%	51.22%	38.36%	9.43%	62.57%	39.34%
35-49, 50+	222	6.31%	46.04%	31.48%	6.76%	41.12%	23.03%
50+, 50+	539	3.34%	17.83%	11.05%	4.82%	16.78%	7.80%
Pair Race							
Hispanic	2,518	2.22%	19.34%	9.80%	2.98%	24.76%	9.22%
black	2,196	2.82%	23.14%	12.07%	2.32%	21.34%	9.83%
white	12,855	1.07%	14.32%	9.63%	1.52%	17.49%	9.60%
other	1,016	3.05%	16.51%	4.36%	7.09%	48.00%	18.21%
white & black	165	4.24%	5.48%	0.97%	8.48%	19.62%	6.91%
white & Hispanic	602	2.99%	29.00%	16.79%	1.66%	28.55%	13.43%
white & other	468	3.21%	20.61%	12.73%	3.21%	23.43%	11.77%
black & Hispanic	62	9.68%	8.71%	2.03%	6.45%	22.08%	10.79%
black & other	70	4.29%	58.72%	28.03%	4.29%	49.97%	10.81%
Hispanic & other	79	3.80%	4.20%	0.74%	12.66%	18.64%	7.53%
Pair Gender							
Male, Male	4,333	2.08%	11.37%	3.82%	2.84%	19.91%	5.88%
Female, Female	4,460	2.04%	10.62%	2.60%	2.04%	12.95%	3.22%
Male, Female	11,238	1.40%	20.26%	13.83%	2.10%	24.26%	13.36%
Household Size							
Two	5,540	0.42%	3.10%	0.62%	0.81%	8.29%	2.83%
Three	5,414	2.03%	15.86%	8.09%	3.18%	23.20%	9.31%
Four or more	9,077	2.26%	23.93%	15.31%	2.57%	27.69%	14.76%

Table J.2 2003 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR03WT10*PR03WT11) Unweighted	Weighted[2]	Outwinsor[3]	After res.pr.nr[1] (SDUWT*PR03WT10*...*PR03WT12) Unweighted	Weighted[2]	Outwinsor[3]
Census Region							
Northeast	3,923	1.89%	17.95%	9.75%	2.42%	25.79%	10.50%
South	5,936	1.35%	14.96%	9.14%	2.04%	18.82%	8.67%
Midwest	5,705	1.67%	19.10%	12.42%	2.09%	20.01%	12.16%
West	4,467	1.99%	16.79%	8.99%	2.57%	23.24%	10.56%
Quarter							
Quarter 1	4,919	1.57%	21.89%	15.64%	1.97%	26.18%	15.78%
Quarter 2	5,039	1.65%	17.30%	8.95%	1.83%	19.93%	9.53%
Quarter 3	5,245	1.66%	13.20%	6.45%	2.25%	19.79%	7.12%
Quarter 4	4,828	1.88%	15.07%	8.73%	2.96%	20.24%	8.75%
% Hispanic in Segment							
50-100%	1,242	2.25%	14.06%	4.96%	3.62%	20.25%	5.79%
10-50%	3,242	2.68%	22.34%	14.26%	2.96%	30.05%	15.14%
<10%	15,547	1.43%	15.67%	9.34%	1.99%	19.11%	9.35%
% Black in Segment							
50-100%	1,569	2.80%	30.23%	18.42%	2.61%	28.87%	13.99%
10-50%	2,709	2.77%	20.89%	11.85%	3.32%	23.89%	10.66%
<10%	15,753	1.39%	14.51%	8.59%	2.03%	20.29%	9.79%
% Owner-Occupied DUs in Segment							
50-100%	15,266	1.56%	17.19%	10.72%	2.12%	21.79%	11.02%
10-50%	3,625	2.51%	15.46%	6.63%	3.09%	20.24%	7.09%
<10%	1,140	0.79%	16.08%	6.83%	1.32%	20.36%	4.50%
Combined Median Rent/Housing Value							
1st Quintile	3,873	1.45%	18.19%	12.34%	1.94%	20.29%	11.75%
2nd Quintile	3,880	1.34%	17.48%	11.85%	1.55%	17.31%	10.54%
3rd Quintile	4,364	1.28%	23.13%	14.82%	2.09%	26.92%	13.82%
4th Quintile	3,869	1.55%	14.23%	7.77%	2.38%	21.65%	8.64%
5th Quintile	4,045	2.82%	11.79%	4.03%	3.26%	19.82%	7.27%
Population Density							
Large MSA	6,996	2.40%	18.37%	9.59%	3.37%	24.13%	10.13%
Medium-Small MSA	7,544	1.46%	16.68%	11.57%	1.75%	20.24%	11.66%
Non-MSA, Urban	2,561	1.25%	12.68%	6.73%	1.41%	17.66%	8.03%
Non-MSA, Rural	2,930	0.96%	15.00%	9.37%	1.57%	17.10%	8.56%
Group Quarters							
Group	483	1.86%	7.58%	1.88%	1.24%	10.32%	0.92%
Nongroup	19,548	1.68%	16.92%	10.00%	2.27%	21.55%	10.28%

[1] This step used demographic variables from screener data for all responding person pairs; Res = respondent, PR = pair, NR = nonresponse adjustment.
[2] Weighted extreme value proportion: 100 * $\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.
[3] Outwinsor weight proportion: 100 * $\sum_k(w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table J.3 2003 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12)			After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14)		
		Unweighted	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**
Total	20,031	2.20%	14.36%	4.64%	1.84%	7.14%	0.69%	0.12%	0.90%	0.03%
Pair Age Group										
12-17, 12-17	3,860	0.13%	0.83%	0.11%	0.39%	2.43%	0.46%	0.00%	0.00%	0.00%
12-17, 18-25	2,672	0.94%	4.52%	1.34%	1.27%	4.29%	0.37%	0.00%	0.00%	0.00%
12-17, 26-34	758	0.53%	2.14%	1.09%	0.26%	1.14%	0.14%	0.13%	1.17%	0.07%
12-17, 35-49	3,180	0.91%	3.93%	0.98%	0.79%	5.42%	0.64%	0.09%	0.61%	0.04%
12-17, 50+	446	1.79%	3.87%	0.98%	0.22%	0.16%	0.02%	0.00%	0.00%	0.00%
18-25, 18-25	4,285	2.87%	14.53%	3.74%	2.54%	8.50%	0.93%	0.00%	0.00%	0.00%
18-25, 26-34	848	2.59%	13.80%	5.83%	2.83%	8.34%	0.93%	0.35%	1.46%	0.03%
18-25, 35-49	1,079	5.28%	14.89%	2.93%	5.28%	13.80%	1.16%	0.28%	0.51%	0.01%
18-25, 50+	453	2.65%	11.39%	2.80%	2.43%	7.39%	0.65%	0.22%	0.62%	0.02%
26-34, 26-34	623	4.17%	23.24%	8.95%	4.49%	12.62%	1.47%	0.16%	0.18%	0.01%
26-34, 35-49	372	10.75%	34.24%	13.94%	3.49%	4.10%	0.37%	0.27%	0.60%	0.02%
26-34, 50+	113	2.65%	5.47%	0.46%	5.31%	11.70%	1.01%	0.88%	2.77%	0.14%
35-49, 35-49	579	7.25%	27.03%	11.73%	2.25%	2.03%	0.39%	0.35%	0.16%	0.01%
35-49, 50+	228	7.02%	29.31%	10.44%	5.26%	10.88%	0.81%	1.32%	2.58%	0.09%
50+, 50+	535	5.42%	15.20%	3.11%	3.36%	8.84%	0.69%	0.93%	1.98%	0.05%
Pair Race										
Hispanic	2,548	2.75%	11.71%	3.16%	1.57%	3.73%	0.33%	0.00%	0.00%	0.00%
black	2,184	2.20%	9.11%	2.42%	1.42%	5.06%	0.59%	0.05%	0.07%	0.00%
white	12,657	1.47%	12.24%	4.03%	1.43%	5.87%	0.51%	0.03%	0.31%	0.01%
other	977	7.98%	51.05%	18.60%	7.37%	31.55%	3.44%	1.74%	10.46%	0.44%
white & black	171	6.43%	13.83%	4.49%	4.09%	15.28%	1.72%	0.00%	0.00%	0.00%
white & Hispanic	595	1.85%	9.41%	1.70%	2.35%	5.55%	0.67%	0.34%	1.31%	0.06%
white & other	607	3.62%	17.76%	5.68%	1.81%	4.79%	0.74%	0.00%	0.00%	0.00%
black & Hispanic	65	4.62%	12.83%	5.83%	7.69%	3.05%	0.33%	0.00%	0.00%	0.00%
black & other	125	4.00%	13.90%	5.23%	1.60%	17.06%	0.80%	0.00%	0.00%	0.00%
Hispanic & other	102	6.86%	13.32%	5.81%	4.90%	2.13%	0.29%	0.00%	0.00%	0.00%
Pair Gender										
Male, Male	4,325	2.96%	14.82%	4.28%	2.36%	8.06%	0.49%	0.05%	0.21%	0.01%
Female, Female	4,458	2.02%	13.16%	3.50%	1.95%	11.52%	1.28%	0.27%	2.30%	0.09%
Male, Female	11,248	1.98%	14.56%	5.05%	1.59%	5.71%	0.59%	0.09%	0.72%	0.03%
Household Size										
Two	5,540	0.88%	9.83%	3.16%	0.74%	4.88%	0.54%	0.16%	1.41%	0.06%
Three	5,414	3.25%	21.63%	7.41%	2.62%	10.18%	0.96%	0.17%	1.19%	0.05%
Four or more	9,077	2.38%	12.82%	3.91%	2.04%	6.70%	0.63%	0.07%	0.47%	0.01%

Table J.3 2003 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

		Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12)			After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14)		
Domain	***n***	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**	**Unweighted**	**Weighted[2]**	**Outwinsor[3]**
Census Region										
Northeast	3,923	2.32%	13.37%	3.71%	2.27%	9.06%	0.62%	0.28%	2.11%	0.07%
South	5,936	1.94%	8.45%	1.97%	1.52%	5.62%	0.38%	0.10%	0.75%	0.03%
Midwest	5,705	2.03%	17.08%	6.19%	1.72%	6.06%	0.74%	0.02%	0.04%	0.00%
West	4,467	2.66%	20.76%	7.61%	2.04%	8.71%	1.13%	0.13%	0.95%	0.04%
Quarter										
Quarter 1	4,919	1.83%	18.05%	8.40%	2.20%	8.27%	0.73%	0.10%	1.13%	0.04%
Quarter 2	5,039	1.61%	11.15%	3.08%	1.49%	5.81%	0.47%	0.14%	1.05%	0.05%
Quarter 3	5,245	2.36%	15.32%	4.33%	1.70%	6.75%	0.66%	0.13%	0.83%	0.03%
Quarter 4	4,828	3.02%	12.96%	2.82%	1.99%	7.77%	0.92%	0.10%	0.60%	0.02%
% Hispanic in Segment										
50-100%	1,242	3.38%	15.90%	3.82%	2.01%	6.19%	0.62%	0.08%	0.60%	0.02%
10-50%	3,242	2.90%	21.68%	8.44%	2.53%	9.83%	1.10%	0.15%	1.35%	0.05%
<10%	15,547	1.96%	11.98%	3.61%	1.68%	6.46%	0.58%	0.12%	0.81%	0.03%
% Black in Segment										
50-100%	1,569	2.10%	6.89%	2.26%	1.72%	6.00%	0.80%	0.06%	0.10%	0.00%
10-50%	2,709	3.06%	15.42%	3.92%	2.55%	9.79%	0.70%	0.18%	0.60%	0.02%
<10%	15,753	2.06%	14.92%	5.02%	1.73%	6.74%	0.68%	0.11%	1.04%	0.04%
% Owner-Occupied DUs in Segment										
50-100%	15,266	2.04%	14.22%	4.75%	1.68%	6.12%	0.61%	0.11%	0.84%	0.03%
10-50%	3,625	3.17%	15.80%	4.33%	2.81%	12.33%	1.11%	0.17%	1.23%	0.04%
<10%	1,140	1.23%	7.06%	2.27%	0.79%	5.10%	0.63%	0.09%	0.76%	0.01%
Combined Median Rent/Housing Value										
1st Quintile	3,873	1.94%	14.92%	6.81%	1.52%	5.74%	0.54%	0.03%	0.26%	0.01%
2nd Quintile	3,880	1.75%	7.46%	2.07%	1.98%	6.68%	0.70%	0.15%	0.63%	0.04%
3rd Quintile	4,364	1.86%	17.66%	6.14%	1.88%	7.78%	0.71%	0.14%	1.46%	0.04%
4th Quintile	3,869	2.20%	14.52%	4.08%	1.63%	6.83%	0.74%	0.18%	1.00%	0.04%
5th Quintile	4,045	3.26%	15.55%	4.21%	2.15%	8.06%	0.73%	0.10%	0.85%	0.04%
Population Density										
Large MSA	6,996	3.44%	17.59%	4.68%	2.22%	9.12%	0.85%	0.24%	1.53%	0.05%
Medium-Small MSA	7,544	1.62%	14.62%	6.30%	1.88%	5.98%	0.58%	0.07%	0.33%	0.02%
Non-MSA, Urban	2,561	1.17%	4.73%	1.40%	1.37%	5.47%	0.68%	0.08%	0.81%	0.04%
Non-MSA, Rural	2,930	1.64%	7.62%	2.33%	1.23%	3.54%	0.40%	0.00%	0.00%	0.00%
Group Quarters										
Group	483	1.24%	10.32%	1.12%	2.90%	13.11%	2.47%	0.00%	0.00%	0.00%
Nongroup	19,548	2.23%	14.38%	4.65%	1.81%	7.12%	0.69%	0.12%	0.90%	0.03%

Table J.3 2003 NSDUH Respondent Pair–Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12) Unweighted	Weighted[2]	Outwinsor[3]	After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13) Unweighted	Weighted[2]	Outwinsor[3]	Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14) Unweighted	Weighted[2]	Outwinsor[3]
Pair Relationship[4]										
Parent-child (12-14)	2,299	0.52%	3.07%	0.83%	0.65%	4.09%	0.39%	0.13%	0.89%	0.06%
Parent-child (12-17)	4,045	0.72%	3.09%	0.82%	0.72%	4.21%	0.50%	0.10%	0.61%	0.04%
Parent-child (12-20)	4,791	1.48%	7.05%	1.55%	1.19%	6.14%	0.62%	0.13%	0.63%	0.03%
Sibling (12-14) - sibling (15-17)	2,250	0.13%	1.02%	0.12%	0.31%	2.22%	0.47%	0.00%	0.00%	0.00%
Sibling (12-17) - sibling (18-25)	2,311	0.78%	3.59%	1.05%	1.13%	3.94%	0.32%	0.00%	0.00%	0.00%
Spouse-spouse	4,169	2.40%	20.67%	7.96%	2.45%	5.44%	0.56%	0.14%	0.52%	0.02%
Spouse-spouse with children (under 18)	2,056	2.19%	28.08%	13.68%	2.19%	3.80%	0.42%	0.19%	0.34%	0.01%

[1] This step used demographic variables from questionnaire data for all responding person pairs; Res = respondent, PR = pair, PS = poststratification adjustment, EV = extreme value adjustment.

[2] Weighted extreme value proportion: $100 * \sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values and w_k denotes the weight for both extreme values and non-extreme values.

[3] Outwinsor weight proportion: $100 * \sum_k (w_{ek}-b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain. Spouse-spouse pair relationships also included partner-partner relationships.

Appendix K: Evaluation of Calibration Weights: Pair-Level Slippage Rates

(This page intentionally left blank.)

Table K.1 2003 NSDUH Respondent Pair-Level Slippage Rates

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control Total from SDU (C)	(I-C)/C %	(F-C)/C %
Total	20,031	198,686,226	198,686,226	198,686,226	0.00	-0.00
Pair Age Group						
12-17, 12-17	3,860	7,270,758	7,287,395	7,287,395	-0.23	0.00
12-17, 18-25	2,672	7,527,252	7,535,408	7,535,408	-0.11	-0.00
12-17, 26-34	758	5,430,537	5,551,816	5,551,816	-2.18	-0.00
12-17, 35-49	3,180	32,022,621	31,882,494	31,882,494	0.44	0.00
12-17, 50+	446	8,784,962	8,854,946	8,854,946	-0.79	0.00
18-25, 18-25	4,285	10,963,014	11,171,886	11,171,886	-1.87	0.00
18-25, 26-34	848	7,366,351	7,079,298	7,079,298	4.05	-0.00
18-25, 35-49	1,079	17,061,357	17,240,371	17,240,371	-1.04	-0.00
18-25, 50+	453	12,105,630	12,061,703	12,061,703	0.36	-0.00
26-34, 26-34	623	9,352,244	10,045,381	10,045,381	-6.90	-0.00
26-34, 35-49	372	8,673,337	8,462,249	8,462,249	2.49	0.00
26-34, 50+	113	7,988,171	7,894,546	7,894,546	1.19	-0.00
35-49, 35-49	579	20,939,549	20,346,680	20,346,680	2.91	-0.00
35-49, 50+	228	14,789,144	14,564,006	14,564,006	1.55	-0.00
50+, 50+	535	28,411,298	28,708,047	28,708,047	-1.03	-0.00
Pair Race						
Hispanic	2,548	29,801,907	29,672,621	29,672,621	0.44	-0.00
black	2,184	22,010,390	21,913,729	21,913,729	0.44	-0.00
white	12,657	118,949,593	121,080,377	121,080,377	-1.76	-0.00
other	977	12,648,243	12,608,039	12,608,039	0.32	-0.00
white & black	171	1,525,433	1,579,917	1,579,917	-3.45	-0.00
white & Hispanic	595	6,298,347	5,778,230	5,778,230	9.00	0.00
white & other	607	5,148,122	4,107,320	4,107,320	25.34	0.00
black & Hispanic	65	823,545	805,621	805,621	2.22	-0.00
black & other	125	748,042	547,620	547,620	36.60	-0.00
Hispanic & other	102	732,604	592,751	592,751	23.59	0.00
Pair Gender						
Male, Male	4,325	35,635,744	35,736,278	35,736,278	-0.28	-0.00
Female, Female	4,458	34,729,798	34,641,869	34,641,869	0.25	-0.00
Male, Female	11,248	128,320,684	128,308,079	128,308,079	0.01	-0.00
Pair Domain[3,4,5]						
Parent-child (12-14) *	2,299	12,304,361	12,753,776	12,753,776	-3.52	0.00
Parent-child (12-17)*	4,045	23,711,852	24,601,628	24,601,628	-3.62	0.00
Parent-child (15-17)*	1,746	11,407,491	11,847,853	11,847,853	-3.72	0.00
Parent-child (12-20)*	4,791	31,905,590	32,508,858	32,508,858	-1.86	-0.00
Parent*-child (12-14)	2,299	19,320,985	19,533,161	19,533,161	-1.09	0.00
Parent*-child (12-17)	4,045	30,469,767	31,683,979	31,683,979	-3.83	0.00
Parent*-child (15-17)	1,746	17,415,719	18,057,247	18,189,079	-4.25	-0.72
Parent*-child (12-20)	4,791	38,667,337	38,443,971	38,443,971	0.58	-0.00
Sibling (12-14) -sibling (15-17)*	2,250	3,855,708	3,926,182	3,926,182	-1.79	0.00
Sibling (12-17) -sibling (18-25)*	2,311	5,263,216	5,208,883	5,208,883	1.04	-0.00
Spouse-spouse/partner-partner	4,169	65,799,937	66,120,938	66,120,938	-0.49	-0.00
Spouse-spouse/partner-partner with children under 18	2,056	29,818,546	29,242,559	29,242,559	1.97	-0.00

Table K.1 2003 NSDUH Respondent Pair-Level Slippage Rates (continued)

Domain	*n*	Initial Total (I)[1]	Final Total (F)[2]	Control Total from SDU (C)	(I-C)/C %	(F-C)/C %
Household Size						
Two	5,540	51,289,036	51,289,036	51,289,036	0.00	-0.00
Three	5,414	52,184,673	52,184,673	52,184,673	0.00	-0.00
Four or more	9,077	95,212,516	95,212,516	95,212,516	0.00	-0.00
Census Region						
Northeast	3,923	37,310,230	37,310,230	37,310,230	0.00	-0.00
South	5,936	67,999,140	67,999,140	67,999,140	0.00	-0.00
Midwest	5,705	43,012,086	43,012,085	43,012,085	0.00	-0.00
West	4,467	50,364,771	50,364,771	50,364,771	0.00	-0.00
Quarter						
Quarter 1	4,919	48,734,003	48,734,003	48,734,003	0.00	-0.00
Quarter 2	5,039	49,464,530	49,464,530	49,464,530	0.00	-0.00
Quarter 3	5,245	50,689,624	50,689,624	50,689,624	0.00	-0.00
Quarter 4	4,828	49,798,069	49,798,069	49,798,069	0.00	-0.00
% Hispanic in Segment						
50-100%	1,242	17,739,499	17,739,499	17,739,499	0.00	-0.00
10-50%	3,242	41,669,379	41,669,379	41,669,379	0.00	-0.00
<10%	15,547	139,277,348	139,277,348	139,277,348	0.00	-0.00
% Black in Segment						
50-100%	1,569	15,619,216	15,619,216	15,619,216	0.00	0.00
10-50%	2,709	30,043,582	30,043,582	30,043,582	0.00	-0.00
<10%	15,753	153,023,428	153,023,427	153,023,428	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	15,266	161,980,266	161,980,265	161,980,266	0.00	-0.00
10-50%	3,625	33,269,477	33,269,477	33,269,477	0.00	-0.00
<10%	1,140	3,436,484	3,436,484	3,436,484	0.00	0.00
Combined Median						
Rent/Housing Value						
1st Quintile	3,873	28,644,752	28,644,752	28,644,752	0.00	-0.00
2nd Quintile	3,880	32,550,207	32,550,207	32,550,207	0.00	-0.00
3rd Quintile	4,364	44,623,622	44,623,622	44,623,622	0.00	-0.00
4th Quintile	3,869	47,341,588	47,341,588	47,341,588	0.00	-0.00
5th Quintile	4,045	45,526,057	45,526,057	45,526,057	0.00	-0.00
Population Density						
Large MSA	6,996	94,790,959	94,790,959	94,790,959	0.00	-0.00
Medium-Small MSA	7,544	63,434,377	63,434,377	63,434,377	0.00	-0.00
Non-MSA, Urban	2,561	16,919,812	16,919,812	16,919,812	0.00	0.00
Non-MSA, Rural	2,930	23,541,079	23,541,079	23,541,079	0.00	-0.00
Group Quarters						
Group	483	685,528	685,528	685,528	0.00	0.00
Nongroup	19,548	198,000,698	198,000,697	198,000,697	0.00	-0.00

[1] YR03WT1*...*YR03WT9*PR03WT10*...*PR03WT12 (before person pair poststratification).

[2] YR03WT1*...*YR03WT9*PR03WT10*...*PR03WT13 (after person pair poststratification).

[3] The member of pair that is the focus is designated with an *.

[4] The parent-child (15-17) pair domains were not controlled for within the modeling and thus have higher slippage rates than the other domains listed. However , since these domains are a subset of other controlled domains, the rates are not large.

[5] Slippage rates were not calculated for the sibling-sibling domains with the younger child as the focus since no household counts for this domain were calculated and are required to construct the appropriate controls totals.

Appendix L: Evaluation of Calibration Weights: Pair-Level Weight Summary Statistics

(This page intentionally left blank)

Table L.1 2003 NSDUH Selected Pair–Level Weight Summary Statistics

Domain	n	Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)						Before sel.pr.ps[1] (SDUWT*PR03WT10)						After sel.pr.ps[1] (SDUWT*PR03WT10*PR03WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	25,447	20	503	756	1,191	8,669	1.46	29	1,081	2,779	7,685	7,860,602	87.87	14	1,022	2,661	7,603	1,721,635	13.55
Pair Age Group																			
12-17, 12-17	4,302	23	439	671	1,078	7,739	1.49	47	683	1,079	1,979	86,414	3.21	14	537	1,091	2,190	20,246	2.29
12-17, 18-25	3,206	22	537	790	1,235	5,437	1.43	29	833	1,406	2,663	46,613	2.84	23	855	1,654	2,918	25,717	2.03
12-17, 26-34	927	47	513	727	1,161	6,121	1.50	223	2,551	4,116	7,119	165,055	3.00	148	1,608	3,581	7,265	120,188	3.27
12-17, 35-49	3,892	38	463	701	1,136	6,174	1.48	197	3,219	5,758	9,660	187,175	2.30	144	2,418	5,385	10,306	84,332	2.31
12-17, 50+	584	46	532	737	1,181	6,690	1.44	566	6,837	10,950	16,799	139,289	2.04	420	5,598	10,740	18,994	159,506	2.04
18-25, 18-25	5,522	20	489	798	1,199	5,263	1.43	42	698	1,231	2,188	87,281	3.40	23	548	1,185	2,548	23,932	2.42
18-25, 26-34	1,085	37	495	809	1,294	4,312	1.47	296	2,559	4,460	7,309	113,980	2.51	157	1,637	3,467	7,387	135,176	3.14
18-25, 35-49	1,533	44	565	804	1,265	5,587	1.41	278	3,984	6,794	12,037	131,997	2.39	186	3,349	7,058	13,964	128,842	2.32
18-25, 50+	695	30	613	917	1,341	8,669	1.46	446	7,814	11,493	19,593	459,725	2.88	190	5,402	11,801	21,925	155,250	2.11
26-34, 26-34	856	59	561	760	1,224	4,291	1.45	403	4,946	7,381	12,170	590,070	7.57	321	3,379	6,890	12,161	533,756	6.28
26-34, 35-49	493	78	527	745	1,259	3,565	1.43	770	5,596	9,040	15,440	1,282,587	19.29	483	3,362	8,194	16,059	916,448	10.37
26-34, 50+	195	73	582	879	1,447	4,213	1.46	1,628	12,525	20,196	34,217	1,468,435	10.14	705	10,983	24,432	39,758	703,542	3.84
35-49, 35-49	892	57	509	741	1,197	4,576	1.43	673	6,185	9,912	15,381	1,839,013	17.92	418	5,786	10,255	18,388	1,721,635	12.87
35-49, 50+	392	30	548	858	1,224	4,322	1.41	487	10,919	17,779	28,070	2,316,587	17.25	309	9,978	17,822	38,040	1,551,582	7.66
50+, 50+	873	61	531	741	1,180	3,469	1.38	1,286	14,195	20,467	31,198	7,860,602	65.81	1,953	14,031	22,800	35,833	1,311,344	5.60
Pair Race																			
Hispanic	3,271	33	516	893	1,504	7,739	1.43	50	1,356	3,183	8,331	2,531,076	31.87	23	1,053	3,123	8,204	916,448	10.23
black	2,669	24	617	859	1,274	5,131	1.36	29	1,215	3,056	7,921	1,468,435	19.33	14	1,045	3,046	7,950	617,849	9.20
white	16,306	29	521	734	1,115	4,291	1.40	76	1,061	2,755	7,613	7,860,602	127.15	32	1,057	2,587	7,474	1,721,635	16.53
other	1,387	20	259	537	1,327	5,448	1.88	48	797	2,238	7,211	445,245	7.14	23	669	2,231	7,599	401,848	6.94
white & black	202	25	553	813	1,256	3,170	1.42	43	1,139	2,453	7,908	65,206	3.21	66	1,263	2,788	8,741	54,498	2.98
white & Hispanic	753	22	454	777	1,279	4,949	1.49	29	1,096	2,616	7,659	693,399	20.68	23	965	2,619	7,659	533,756	11.57
white & other	593	29	264	528	988	5,018	1.79	71	853	2,302	6,338	1,244,120	44.46	53	748	2,233	5,950	454,254	11.17
black & Hispanic	80	23	385	995	1,689	8,669	2.19	67	1,130	2,408	10,481	459,725	17.81	179	984	3,705	11,132	97,212	3.81
black & other	89	53	463	718	1,195	4,312	1.69	188	886	2,638	7,278	1,282,587	48.67	60	451	1,420	3,750	203,258	13.77
Hispanic & other	97	46	214	345	999	3,823	2.09	67	562	1,521	5,351	32,172	3.15	55	647	2,092	6,594	57,992	3.48
Pair Gender																			
Male, Male	5,568	30	509	796	1,220	8,669	1.46	43	1,013	2,426	6,573	1,282,587	15.08	23	1,010	2,463	6,359	276,533	5.56
Female, Female	5,347	20	472	723	1,164	5,919	1.48	29	1,035	2,504	6,467	187,175	4.04	23	937	2,378	6,416	247,119	5.10
Male, Female	14,532	22	511	756	1,189	7,739	1.44	29	1,133	3,123	8,590	7,860,602	108.09	14	1,053	2,928	8,603	1,721,635	16.42
Household Size																			
Two	7,392	20	477	731	1,144	4,402	1.40	29	1,013	2,987	9,292	118,947	2.71	23	729	2,074	8,276	154,183	3.70
Three	6,947	22	511	754	1,182	6,174	1.46	29	1,208	3,352	6,935	7,272,546	154.34	14	1,222	3,311	7,445	1,721,635	16.84
Four or more	11,108	23	516	776	1,241	8,669	1.48	49	1,068	2,449	7,448	7,860,602	88.47	23	1,157	2,593	7,236	1,551,582	16.05

Table L.1 2003 NSDUH Selected Pair–Level Weight Summary Statistics (continued)

Domain	*n*	Screener DU-Level Weights (SDUWT: YR03WT1*...*YR03WT9)						Before sel.pr.ps[1] (SDUWT*PR03WT10)						After sel.pr.ps[1] (SDUWT*PR03WT10*PR03WT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region																			
Northeast	5,097	23	369	746	1,032	6,174	1.45	29	1,006	2,309	7,359	7,860,602	177.68	31	976	2,535	7,251	1,049,307	12.09
South	7,531	20	673	1,001	1,362	5,744	1.32	47	1,485	3,764	9,624	2,316,587	16.79	14	1,379	3,783	9,603	926,052	8.01
Midwest	7,123	41	537	643	831	5,018	1.28	42	905	2,493	6,554	2,531,076	44.63	27	981	2,295	5,751	1,551,582	18.81
West	5,696	22	280	670	1,621	8,669	1.67	29	1,016	2,621	8,019	7,272,546	120.44	23	757	2,379	7,525	1,721,635	17.77
Quarter																			
Quarter1	6,237	24	525	785	1,211	7,739	1.46	29	1,067	2,698	7,706	2,531,076	34.07	19	964	2,478	7,528	1,721,635	20.66
Quarter2	6,345	22	490	743	1,181	8,669	1.47	29	1,081	2,895	7,885	7,860,602	131.71	16	959	2,633	7,467	1,049,307	10.81
Quarter3	6,736	23	494	723	1,157	6,690	1.45	48	1,078	2,707	7,370	1,282,587	16.04	23	1,085	2,791	7,666	916,448	9.61
Quarter4	6,129	20	503	775	1,228	6,121	1.44	48	1,107	2,796	7,787	7,272,546	133.82	14	1,073	2,770	7,688	1,311,344	13.15
% Hispanic in Segment																			
50-100%	1,605	79	693	1,198	1,664	7,739	1.28	115	1,927	4,533	10,020	1,062,759	11.01	53	1,646	4,349	10,345	916,448	8.36
10-50%	4,222	30	561	1,000	1,558	8,669	1.41	47	1,414	3,490	9,852	1,839,013	15.67	19	1,114	3,270	9,523	1,721,635	13.56
<10%	19,620	20	470	703	1,073	5,919	1.45	29	994	2,549	7,142	7,860,602	122.55	14	977	2,496	7,063	1,551,582	13.96
% Black in Segment																			
50-100%	1,914	20	566	808	1,205	6,174	1.46	48	1,095	2,790	7,797	1,468,435	29.89	14	855	2,478	7,234	617,849	11.73
10-50%	3,460	40	627	933	1,378	5,744	1.35	47	1,281	3,084	7,948	1,244,120	15.37	33	1,231	3,216	8,332	902,983	10.00
<10%	20,073	22	462	721	1,157	8,669	1.47	29	1,047	2,722	7,639	7,860,602	108.03	23	1,006	2,585	7,530	1,721,635	14.50
% Owner-Occupied DUs in Segment																			
50-100%	19,452	22	500	737	1,162	8,669	1.45	29	1,111	3,002	8,045	7,860,602	99.61	23	1,146	2,951	8,282	1,721,635	13.71
10-50%	4,573	24	494	819	1,272	7,739	1.48	29	1,015	2,405	7,021	1,062,759	11.36	23	930	2,457	6,654	916,448	10.07
<10%	1,422	20	557	886	1,296	4,963	1.41	47	941	1,848	4,900	505,251	15.65	14	407	881	2,160	202,067	11.05
Combined Median Rent/Housing Value																			
1st Quintile	4,754	38	388	615	851	4,949	1.45	58	887	2,232	5,903	1,229,180	17.34	29	875	2,214	5,815	1,551,582	21.67
2nd Quintile	4,896	22	327	657	1,043	5,263	1.57	29	885	2,195	6,551	2,531,076	57.74	23	853	2,293	6,435	926,052	12.92
3rd Quintile	5,542	33	522	763	1,156	6,121	1.42	50	1,128	2,783	7,693	1,839,013	24.83	23	1,025	2,513	7,260	1,721,635	18.70
4th Quintile	5,047	37	641	996	1,407	8,669	1.36	49	1,403	3,714	9,765	7,272,546	104.84	16	1,268	3,253	9,146	1,311,344	10.83
5th Quintile	5,208	20	552	865	1,340	6,174	1.42	47	1,243	3,229	8,447	7,860,602	151.11	14	1,203	3,383	9,458	1,049,307	7.42
Population Density																			
Large MSA	9,187	20	673	989	1,469	8,669	1.31	47	1,577	4,149	10,358	7,860,602	106.66	14	1,533	3,941	10,447	1,311,344	9.52
Medium-Small MSA	9,409	25	415	686	1,070	7,739	1.46	48	966	2,365	6,633	1,839,013	24.71	31	906	2,335	6,604	1,721,635	21.25
Non-MSA, Urban	3,136	22	262	575	934	4,161	1.60	29	742	1,721	5,353	414,016	9.28	23	622	1,702	5,168	399,666	8.39
Non-MSA, Rural	3,715	30	293	597	933	5,018	1.56	50	865	2,184	6,222	2,316,587	41.18	23	826	2,241	6,233	926,052	12.42
Group Quarters																			
Group	545	24	282	596	1,035	2,967	1.56	29	441	849	1,416	11,262	2.45	36	348	829	1,460	14,447	2.82
Nongroup	24,902	20	508	759	1,194	8,669	1.45	29	1,108	2,894	7,840	7,860,602	86.54	14	1,056	2,759	7,782	1,721,635	13.35

[1] This step used demographic variables from screener data for all selected person pairs; Sel = selected, PR = pair, PS = poststratification.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table L.2 2003 NSDUH Respondent Pair–Level Weight Summary Statistics

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR03WT10*PR03WT11)						After res.pr.nr[1] (SDUWT*PR03WT10*...*PR03WT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,031	14	949	2,440	6,828	1,721,635	14.81	14	1,046	2,813	8,339	2,254,659	14.72
Pair Age Group													
12-17, 12-17	3,870	14	541	1,087	2,214	20,246	2.31	14	577	1,190	2,445	22,778	2.29
12-17, 18-25	2,675	23	852	1,634	2,935	25,717	2.04	23	997	1,931	3,579	43,759	2.17
12-17, 26-34	766	148	1,613	3,539	7,074	120,188	3.50	152	1,776	4,146	8,288	167,782	3.97
12-17, 35-49	3,168	144	2,348	5,328	10,430	84,332	2.30	144	2,602	6,279	12,532	143,091	2.49
12-17, 50+	444	420	6,023	11,008	19,203	159,506	2.04	420	7,122	14,100	25,239	145,093	2.03
18-25, 18-25	4,362	23	536	1,136	2,525	23,932	2.42	23	566	1,266	3,044	51,560	2.93
18-25, 26-34	804	157	1,588	3,342	7,050	93,453	2.83	157	1,738	3,889	9,814	118,435	3.35
18-25, 35-49	1,087	186	3,311	7,038	14,114	128,842	2.40	186	3,737	8,881	19,817	197,873	2.59
18-25, 50+	449	583	5,249	12,359	21,979	132,323	2.06	712	7,351	17,806	34,533	223,128	2.20
26-34, 26-34	600	321	3,359	6,694	11,696	533,756	7.77	321	3,550	7,639	14,648	599,409	7.58
26-34, 35-49	345	540	3,385	8,237	15,916	374,617	5.69	540	3,634	10,280	21,869	590,630	6.13
26-34, 50+	106	1,951	9,644	23,239	39,032	703,542	4.92	2,028	20,755	42,694	84,986	810,153	3.12
35-49, 35-49	594	759	5,610	10,477	18,222	1,721,635	14.91	782	6,341	12,225	21,460	2,254,659	14.96
35-49, 50+	222	309	9,155	16,776	35,778	1,551,582	10.31	310	12,920	22,991	63,667	1,998,204	7.19
50+, 50+	539	2,648	13,505	21,659	33,447	902,983	4.03	2,926	21,609	38,391	64,700	1,085,672	2.76
Pair Race													
Hispanic	2,518	23	1,010	2,926	7,404	411,195	8.09	23	1,257	3,528	9,714	601,230	8.89
black	2,196	14	1,011	2,910	7,516	617,849	10.69	14	1,081	3,275	8,773	823,995	10.97
white	12,855	32	990	2,386	6,663	1,721,635	18.49	32	1,083	2,757	8,147	2,254,659	18.26
other	1,016	23	554	1,675	5,412	160,802	5.94	23	567	1,857	6,803	278,505	8.03
white & black	165	66	1,102	2,258	7,291	54,498	3.13	68	1,272	2,993	11,869	78,338	3.32
White & Hispanic	602	23	825	2,455	7,231	533,756	13.51	23	834	2,538	8,203	533,770	13.94
white & other	468	57	739	2,107	5,709	454,254	13.84	76	771	2,402	7,112	566,188	13.23
black & Hispanic	62	179	866	3,377	7,817	90,360	3.83	186	1,083	3,612	10,372	90,360	3.81
black & other	70	60	417	1,310	2,826	203,258	16.84	60	417	1,392	3,818	250,322	16.10
Hispanic & other	79	55	492	2,046	4,446	40,036	3.39	55	773	2,716	9,938	63,764	3.35
Pair Gender													
Male, Male	4,333	23	947	2,281	5,540	276,533	5.98	23	1,059	2,595	6,980	523,411	7.67
Female, Female	4,460	23	907	2,257	5,927	169,355	4.79	23	943	2,483	6,886	306,268	6.59
Male, Female	11,238	14	978	2,604	7,677	1,721,635	18.29	14	1,078	3,103	9,545	2,254,659	17.05
Household Size													
Two	5,540	23	669	1,770	7,054	154,183	3.86	23	708	1,883	8,054	306,268	6.15
Three	5,414	14	1,092	3,053	6,866	1,721,635	17.94	14	1,278	3,668	8,858	2,254,659	16.05
Four or more	9,077	23	1,105	2,482	6,683	1,551,582	17.25	23	1,238	2,872	8,114	1,998,204	18.04

Table L.2 2003 NSDUH Respondent Pair–Level Weight Summary Statistics (continued)

		Before res.pr.nr[1] (SDUWT*PR03WT10*PR03WT11)						After res.pr.nr[1] (SDUWT*PR03WT10*...*PR03WT12)					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
Census Region													
Northeast	3,923	31	903	2,301	6,298	703,542	10.42	31	1,005	2,758	7,927	841,482	11.25
South	5,936	14	1,268	3,406	8,782	926,052	9.62	14	1,449	4,033	10,763	1,162,890	9.78
Midwest	5,705	27	928	2,129	5,199	1,551,582	24.89	29	996	2,375	6,126	1,998,204	24.42
West	4,467	23	709	2,107	6,605	1,721,635	17.46	23	752	2,380	8,237	2,254,659	16.69
Quarter													
Quarter1	4,919	19	905	2,331	6,716	1,721,635	27.92	19	972	2,580	8,083	2,254,659	27.27
Quarter2	5,039	16	908	2,433	6,787	703,542	10.07	16	997	2,757	8,200	841,482	11.61
Quarter3	5,245	23	1,007	2,555	6,966	902,983	8.81	23	1,103	2,942	8,459	1,085,672	9.04
Quarter4	4,828	14	992	2,502	6,768	926,052	11.45	14	1,110	3,009	8,697	1,162,890	11.25
% Hispanic in Segment													
50-100%	1,242	53	1,645	4,135	9,419	375,109	4.88	54	1,923	5,043	12,639	599,409	6.05
10-50%	3,242	19	1,028	2,882	7,905	1,721,635	18.73	19	1,154	3,336	9,890	2,254,659	17.59
<10%	15,547	14	910	2,314	6,361	1,551,582	14.53	14	995	2,613	7,716	1,998,204	14.47
% Black in Segment													
50-100%	1,569	14	809	2,276	6,706	617,849	14.31	14	856	2,553	7,664	823,995	13.98
10-50%	2,709	33	1,174	3,013	7,359	902,983	12.29	33	1,319	3,546	9,236	1,085,672	11.34
<10%	15,753	23	930	2,370	6,718	1,721,635	15.39	23	1,020	2,733	8,243	2,254,659	15.51
% Owner-Occupied DUs in Segment													
50-100%	15,266	23	1,064	2,665	7,351	1,721,635	15.57	23	1,175	3,107	9,125	2,254,659	15.20
10-50%	3,625	23	864	2,320	6,109	374,617	7.60	23	942	2,613	7,551	590,630	8.92
<10%	1,140	14	382	839	1,956	202,067	13.63	14	409	896	2,233	235,773	15.01
Combined Median Rent/Housing Value													
1st Quintile	3,873	29	824	2,012	5,173	1,551,582	28.11	29	888	2,252	6,292	1,998,204	27.69
2nd Quintile	3,880	23	825	2,109	5,879	926,052	16.53	23	868	2,325	6,983	1,162,890	15.94
3rd Quintile	4,364	23	950	2,293	6,533	1,721,635	22.47	23	1,046	2,613	7,876	2,254,659	21.83
4th Quintile	3,869	16	1,135	2,928	8,153	902,983	8.67	16	1,314	3,579	10,157	1,085,672	9.17
5th Quintile	4,045	14	1,131	3,070	8,635	411,195	4.97	14	1,250	3,648	11,043	841,482	6.86
Population Density													
Large MSA	6,996	14	1,423	3,488	9,411	902,983	8.35	14	1,607	4,190	12,133	1,085,672	8.59
Medium-Small MSA	7,544	31	852	2,125	5,913	1,721,635	26.29	31	918	2,434	6,992	2,254,659	25.24
Non-MSA, Urban	2,561	23	599	1,572	4,501	399,666	9.65	23	631	1,683	5,199	841,482	15.69
Non-MSA, Rural	2,930	23	798	2,066	5,638	926,052	14.85	23	873	2,402	7,183	1,162,890	14.60
Group Quarters													
Group	483	36	346	819	1,440	13,881	2.78	36	348	816	1,489	23,483	3.87
Nongroup	19,548	14	987	2,530	7,008	1,721,635	14.57	14	1,087	2,930	8,544	2,254,659	14.46

1 This step used demographic variables from screener data for all selected person pairs; Res = respondent, PR = pair, NR = nonresponse adjustment.
2 Q1 and Q3 refer to the first and third quartile of the weight distribution.
3 Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.

Table L.3 2003 NSDUH Respondent Pair–Level Weight Summary Statistics

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12)						After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,031	14	1,046	2,813	8,339	2,254,659	14.72	4	886	2,774	8,382	1,126,996	11.28	4	877	2,773	8,409	1,138,771	11.29
Pair Age Group																			
12-17, 12-17	3,860	14	577	1,191	2,444	22,778	2.29	4	487	1,177	2,408	38,134	2.56	4	482	1,174	2,415	24,967	2.49
12-17, 18-25	2,672	23	994	1,938	3,576	43,759	2.17	8	861	1,830	3,548	31,180	2.25	8	855	1,838	3,588	23,381	2.24
12-17, 26-34	758	120	1,713	4,119	8,220	167,782	4.06	58	1,577	4,026	8,453	164,400	3.99	50	1,517	4,024	8,531	151,382	3.93
12-17, 35-49	3,180	144	2,613	6,273	12,558	143,091	2.49	56	2,198	6,135	12,605	215,840	2.77	52	2,170	6,120	12,725	165,022	2.66
12-17, 50+	446	420	6,707	13,734	25,087	145,093	2.06	158	6,192	14,183	25,825	145,386	2.08	142	6,245	13,966	26,304	136,144	2.09
18-25, 18-25	4,285	23	562	1,254	3,031	51,560	2.98	12	389	1,090	3,352	34,108	2.96	12	379	1,096	3,378	23,479	2.91
18-25, 26-34	848	110	1,705	3,792	8,759	224,182	4.29	40	1,243	3,598	9,313	141,678	3.65	36	1,203	3,568	9,615	140,239	3.71
18-25, 35-49	1,079	186	3,677	8,881	19,952	197,873	2.57	77	3,013	8,742	19,911	178,925	2.82	80	3,060	8,946	20,128	165,101	2.73
18-25, 50+	453	712	7,433	17,583	33,905	223,128	2.20	345	6,619	16,930	33,981	201,589	2.30	321	6,734	16,497	34,402	171,762	2.25
26-34, 26-34	623	270	3,353	7,363	14,538	533,770	6.13	114	3,733	7,778	15,997	449,774	5.35	105	3,788	7,964	15,919	451,834	5.41
26-34, 35-49	372	540	3,627	10,307	21,656	590,630	5.95	195	4,322	12,090	23,090	677,486	6.52	201	4,141	11,694	23,545	675,555	6.49
26-34, 50+	113	2,028	19,460	40,225	75,666	810,153	3.25	750	14,169	42,275	86,594	686,437	2.94	741	14,490	41,602	85,139	675,218	2.90
35-49, 35-49	579	782	6,341	12,356	21,460	2,254,659	14.35	317	6,284	11,412	21,795	1,085,329	10.94	286	5,935	11,318	21,244	1,047,515	11.04
35-49, 50+	228	310	11,985	22,987	61,961	1,998,204	7.19	101	11,557	22,359	72,069	1,126,996	4.22	96	11,858	22,383	71,999	1,138,771	4.31
50+, 50+	535	2,926	21,535	38,386	64,568	1,085,672	2.77	1,052	22,948	42,016	68,398	758,943	2.22	965	23,291	42,635	68,490	760,921	2.20
Pair Race																			
Hispanic	2,548	23	1,262	3,522	9,819	601,230	8.84	10	1,113	3,303	9,601	1,085,329	11.66	9	1,087	3,318	9,772	1,047,515	11.40
black	2,184	14	1,044	3,082	8,756	823,995	10.91	4	804	3,113	8,876	758,943	10.65	4	798	3,099	8,832	760,921	10.54
white	12,657	33	1,079	2,753	8,082	2,254,659	18.49	11	986	2,811	8,262	1,126,996	11.39	10	978	2,813	8,257	1,138,771	11.53
other	977	31	655	2,022	7,074	298,887	8.11	10	286	1,433	8,434	450,630	8.52	10	286	1,443	8,582	462,339	8.23
white & black	171	68	1,204	2,993	10,184	78,338	3.40	47	962	2,858	9,555	114,627	4.39	47	993	3,189	9,949	98,631	4.15
white & Hispanic	595	23	843	2,672	9,615	533,770	12.44	8	620	2,118	8,196	449,774	12.88	8	593	2,167	8,122	451,834	13.01
white & other	607	86	800	2,340	7,297	566,188	11.72	31	403	1,509	5,607	518,814	14.43	30	394	1,467	5,678	525,177	14.60
black & Hispanic	65	80	1,186	3,591	7,300	250,322	7.81	51	1,100	3,597	8,401	232,563	7.39	50	1,062	3,526	8,987	235,553	7.56
black & other	125	23	713	1,886	5,639	52,235	3.83	19	399	1,283	4,111	53,838	4.59	19	401	1,309	4,152	48,588	4.37
Hispanic & other	102	80	704	2,384	7,551	66,908	3.80	27	419	1,943	6,483	64,328	4.07	26	393	1,827	6,411	65,143	4.17
Pair Gender																			
Male, Male	4,325	23	1,061	2,596	6,957	523,411	7.68	14	868	2,526	7,179	541,340	7.56	14	861	2,516	7,166	527,926	7.45
Female, Female	4,458	23	951	2,493	6,919	306,268	6.43	10	794	2,287	6,956	452,925	7.05	9	795	2,287	6,891	485,484	6.91
Male, Female	11,248	14	1,075	3,089	9,529	2,254,659	17.09	4	940	3,077	9,513	1,126,996	12.37	4	929	3,071	9,564	1,138,771	12.43
Household Size																			
Two	5,540	23	708	1,883	8,054	306,268	6.15	11	473	1,511	7,917	452,925	6.15	10	465	1,487	7,903	485,484	6.16
Three	5,414	14	1,278	3,668	8,858	2,254,659	16.05	4	1,232	3,845	9,047	712,442	7.54	4	1,223	3,845	9,072	758,408	7.56
Four or more	9,077	23	1,238	2,872	8,114	1,998,204	18.04	8	1,142	2,921	8,074	1,126,996	15.55	8	1,139	2,917	8,103	1,138,771	15.54

Table L.3 2003 NSDUH Respondent Pair–Level Weight Summary Statistics (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12)						After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region																			
Northeast	3,923	31	1,005	2,758	7,927	841,482	11.25	14	947	2,726	7,968	896,618	11.43	14	934	2,712	7,919	893,033	11.35
South	5,936	14	1,449	4,033	10,763	1,162,890	9.78	4	1,342	4,009	10,829	1,126,996	9.89	4	1,325	3,992	10,859	1,138,771	9.94
Midwest	5,705	29	996	2,375	6,126	1,998,204	24.42	12	806	2,179	6,344	712,442	11.98	12	813	2,168	6,306	747,079	12.21
West	4,467	23	752	2,380	8,237	2,254,659	16.69	8	494	2,123	8,077	1,085,329	11.62	8	471	2,113	8,192	1,047,515	11.48
Quarter																			
Quarter1	4,919	19	972	2,580	8,083	2,254,659	27.27	6	886	2,745	8,511	872,597	11.57	5	882	2,742	8,559	880,701	11.71
Quarter2	5,039	16	997	2,757	8,200	841,482	11.61	6	814	2,625	8,065	896,618	12.04	6	807	2,599	8,061	893,033	12.12
Quarter3	5,245	23	1,103	2,942	8,459	1,085,672	9.04	10	949	2,873	8,466	712,442	8.38	10	936	2,859	8,519	727,286	8.43
Quarter4	4,828	14	1,110	3,009	8,697	1,162,890	11.25	4	882	2,841	8,515	1,126,996	13.05	4	873	2,884	8,566	1,138,771	12.82
% Hispanic in Segment																			
50-100%	1,242	54	1,923	5,043	12,639	599,409	6.05	48	1,732	4,803	12,898	1,085,329	9.04	47	1,746	4,991	13,000	1,047,515	8.67
10-50%	3,242	19	1,154	3,336	9,890	2,254,659	17.59	6	1,051	3,379	10,597	712,442	9.56	5	1,032	3,373	10,577	758,408	9.54
<10%	15,547	14	995	2,613	7,716	1,998,204	14.47	4	820	2,521	7,780	1,126,996	11.89	4	814	2,511	7,755	1,138,771	11.98
% Black in Segment																			
50-100%	1,569	14	856	2,553	7,664	823,995	13.98	4	673	2,480	7,620	758,943	13.62	4	663	2,473	7,702	760,921	13.55
10-50%	2,709	33	1,319	3,546	9,236	1,085,672	11.34	12	1,116	3,491	9,521	1,085,329	11.62	11	1,133	3,498	9,587	1,047,515	11.34
<10%	15,753	23	1,020	2,733	8,243	2,254,659	15.51	8	870	2,699	8,263	1,126,996	10.93	8	861	2,679	8,290	1,138,771	11.01
% Owner-Occupied DUs in Segment																			
50-100%	15,266	23	1,175	3,107	9,125	2,254,659	15.20	8	1,065	3,127	9,228	1,126,996	11.10	8	1,054	3,114	9,240	1,138,771	11.12
10-50%	3,625	23	942	2,613	7,551	590,630	8.92	9	732	2,377	7,431	677,486	9.65	9	729	2,359	7,439	675,555	9.53
<10%	1,140	14	409	896	2,233	235,773	15.01	4	244	646	1,878	253,581	21.58	4	233	626	1,898	252,715	21.95
Combined Median Rent/Housing Value																			
1st Quintile	3,873	29	888	2,252	6,292	1,998,204	27.69	10	721	2,138	6,637	852,454	12.71	9	712	2,106	6,586	860,988	13.05
2nd Quintile	3,880	23	868	2,325	6,983	1,162,890	15.94	8	683	2,191	7,115	1,126,996	14.73	8	674	2,203	7,125	1,138,771	14.83
3rd Quintile	4,364	23	1,046	2,613	7,876	2,254,659	21.83	12	880	2,561	7,928	1,085,329	13.87	11	875	2,581	7,884	1,047,515	13.86
4th Quintile	3,869	16	1,314	3,579	10,157	1,085,672	9.17	6	1,068	3,352	9,946	712,442	9.60	6	1,052	3,338	9,938	727,286	9.57
5th Quintile	4,045	14	1,250	3,648	11,043	841,482	6.86	4	1,133	3,677	11,260	896,618	7.10	4	1,132	3,717	11,352	893,033	6.96
Population Density																			
Large MSA	6,996	14	1,607	4,190	12,133	1,085,672	8.59	4	1,426	4,118	12,256	1,085,329	8.72	4	1,421	4,133	12,322	1,047,515	8.62
Medium-Small MSA	7,544	31	918	2,434	6,992	2,254,659	25.24	10	815	2,506	7,369	852,454	10.51	10	815	2,502	7,322	860,988	10.71
Non-MSA, Urban	2,561	23	631	1,683	5,199	841,482	15.69	8	372	1,396	4,711	896,618	19.76	8	363	1,405	4,649	893,033	19.77
Non-MSA, Rural	2,930	23	873	2,402	7,183	1,162,890	14.60	10	734	2,126	7,076	1,126,996	17.05	9	722	2,109	7,006	1,138,771	17.22
Group Quarters																			
Group	483	36	348	816	1,489	23,483	3.87	21	217	633	1,312	21,797	4.66	21	227	610	1,291	23,479	4.63
Nongroup	19,548	14	1,087	2,930	8,544	2,254,659	14.46	4	935	2,883	8,641	1,126,996	11.09	4	923	2,889	8,652	1,138,771	11.09

Table L.3 2003 NSDUH Respondent Pair-Level Weight Summary Statistics (continued)

		Before res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT12)						After res.pr.ps[1] (SDUWT*PR03WT10*...*PR03WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR03WT10*...*PR03WT14)					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
***Pair Relationship*[4]**																			
Parent-child (12-14)	2,299	107	2,305	5,476	11,495	167,782	2.81	56	2,007	5,565	11,624	196,357	3.01	50	1,969	5,489	11,668	161,244	2.94
Parent-child (12-17)	4,045	107	2,552	6,081	12,607	167,782	2.73	56	2,256	6,152	13,005	215,840	2.94	50	2,198	6,123	13,075	165,022	2.85
Parent-child (12-20)	4,791	107	2,758	6,682	13,995	223,128	2.96	56	2,384	6,566	14,153	215,840	3.13	50	2,377	6,586	14,388	168,440	3.03
Sibling (12-14) - sibling (15-17)	2,250	14	603	1,190	2,404	22,778	2.27	4	544	1,215	2,398	38,134	2.49	4	546	1,210	2,420	24,967	2.41
Sibling (12-17) - sibling (18-25)	2,311	23	1,020	1,920	3,481	38,913	2.10	12	884	1,814	3,422	31,180	2.22	11	885	1,812	3,473	23,381	2.21
Spouse-spouse	4,169	23	885	3,276	12,433	2,254,659	18.67	12	764	3,408	13,473	1,085,329	11.76	12	756	3,412	13,272	1,047,515	11.88
Spouse-spouse with children (under 18)	2,056	57	897	2,981	9,168	2,254,659	36.04	28	849	3,321	10,238	1,085,329	21.56	25	818	3,262	10,213	1,047,515	21.84

[1] This step used demographic variables from questionnaire data for all selected person pairs; Res = respondent, PR = pair, PS = poststratification adjustment, EV = extreme value adjustment.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect defined as $1+[(n-1)/n]*CV^2$, where CV=coefficient of variation of weights.
[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain. Spouse-spouse pair relationships also included partner-partner relationships.

Appendix M: Hot-Deck Method of Imputation

Appendix M: Hot-Deck Method of Imputation

M.1 Introduction

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

Three different methods of hot-deck imputations are discussed in this document, though only two were used in the 2003 National Survey on Drug Use and Health (NSDUH)[17]: unweighted sequential hot deck, unweighted random nearest neighbor hot deck (NNHD), and weighted sequential hot deck. The first method, the unweighted sequential hot deck, was the exclusive method of hot-deck imputation used for the 1991 to 1998 surveys and the paper-and-pencil interviewing (PAPI) sample of the 1999 survey. This method was used for all demographic variables in the 1999 survey, but no other variables. In the 2000 NSDUH, the unweighted sequential hot deck method was only used for education and employment status, and was not used at all in 2001 or 2002 surveys. However, it remains in this appendix for historical purposes and for the sake of comparison with the other two methods. The imputation of demographic and other variables unrelated to pair analyses are described in the NSDUH imputation report (Grau et al., 2005). In a similar manner to the 1999 (computer-assisted interviewing [CAI] sample of the survey), 2000, and 2001 surveys, the 2002 and 2003 NSDUHs primarily used the second hot-deck method listed, the unweighted random NNHD. The third hot-deck method, weighted sequential hot deck, incorporated the sampling weights associated with each respondent. Starting in the 2002 NSDUH, the immigrant variable imputations described in the NSDUH imputation report utilized the weighted sequential hot-deck method. For more information on weighted sequential hot-deck see Cox (1980, pp. 721-725) and Iannacchione (1982).

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

M.2 Formation of Imputation Classes

When there was a strong logical association between the base variable and certain auxiliary variables, the dataset was partitioned by the auxiliary variables and imputation

[17] This report presents information from the 2003 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

procedures were implemented independently within classes defined by the cross of the auxiliary variables. These classes were defined by logical and likeness constraints, which are described in the main body of this report. Classes defined by the likeness constraints were collapsed if insufficient donors were available, and classes defined by logical constraints were not collapsed, due to the possibility of an inconsistency with pre-existing nonmissing values that would have resulted.

M.3 Unweighted Sequential Hot Deck

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

M.3.1 Sorting the File

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

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

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

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

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

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

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

M.3.2 Replacing Missing Values

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

M.3.3 Potential Problem

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

M.4 Unweighted Random Nearest Neighbor Hot Deck

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

M.4.1 Creating a Neighborhood of Potential Donors

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

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

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

M.5 Weighted Sequential Hot Deck

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

M.5.1 Sorting the File

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

M.5.2 Replacing Missing Values

The procedure used in the 2003 NSDUH followed directly from Cox (1980). Specifically, once the imputation classes are formed, the data is divided into two data sets: one for respondent and one for nonrespondents. Scaled weights $v(j)$ are then derived for all nonrespondents using the following formula:

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

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

This selection algorithm is an adaptation of Chromy's (1979) sequential sample selection method, which could be implemented using the Chromy-Williams sample selection software (Williams and Chromy, 1980). Furthermore, Iannacchione (1982) revised the Chromy-Williams sample selection software, so that each step of the weighted sequential hot deck is executed in one macro run.

M.5.3 Benefits of Weighted Sequential Hot-Deck

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

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

Appendix N: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

Appendix N: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

N.1 Introduction

The 2003 National Survey on Drug Use and Health (NSDUH)[18] used a predictive mean neighborhood (PMN) method for imputing missing values. This method was implemented in the past several surveys. Starting with the 1999 survey, this PMN method was a new approach, which was developed for the imputation of missing values in the computer-assisted interviewing (CAI) sample. This approach has been used since the 1999 NSDUH[19] and can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, it incorporates predictive means from models and the assignment of imputed values using neighborhoods determined by those predictive means.

N.2 Overview

N.2.1 Predictive Mean Neighborhoods: Derived from Combining Nearest Neighbor Hot Deck and Predictive Mean Matching

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

A commonly used imputation method is a random NNHD (Little & Rubin, 1987, p. 65). With this method, donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (the donor has complete data, the recipient does not). A donor set deemed close to the recipient with respect to a number of covariates is used to select a donor at random. For the NSDUH, the set of covariates typically included demographic variables, as well as some other nonmissing drug use variables. In the case of the NSDUH, to further ensure that a donor matched the recipient as closely as possible, discrete variables (or discrete categories of continuous variables) strongly correlated with the response variable(s) of interest were often used to restrict the set of donors. Furthermore, other restrictions involving outcome variables were imposed on the neighborhood.

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

---

[18] This report presents information from the 2003 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years old or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

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

The PMM method is only applicable to continuous outcome variables. With this method, a distance function is used to determine distances between the predictive mean for the recipient, obtained under a model, and the response variable outcomes for candidate donors. The respondent with the smallest distance is chosen as the donor. Unlike the NNHD, the donor is not randomly selected from a neighborhood. The advantages of PMM include the following:

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

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

N.2.2 Univariate and Multivariate Applications of Predictive Mean Neighborhoods

PMN method is easily applicable to problems of both univariate and multivariate imputations. The need for univariate imputation arises when the value of a single variable, which cannot be easily grouped together with other variables, is missing for the respondent. On the other hand, the need for multivariate imputation arises when values of two or more related variables are missing for a single respondent. The case of a single polytomous variable with missing values can also be viewed as a multivariate imputation problem. An example of this in pair applications is a missing pair relationship for a pair where both respondents are in the 21 to 25-year-old age range. In this instance, the possible outcomes are spouse-spouse without children, spouse-spouse with children, and all other pair relationships.

The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be tedious in practice because of the computational problems due to the volume of model parameters, and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, none of the existing solutions takes the survey design into account because of the obvious problem of specifying the probability distribution underlying survey data. However, in the application of the multivariate predictive mean neighborhood (MPMN) imputation to the 2003 survey, a multivariate model was fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the hierarchy had a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for use of univariate techniques for fitting multivariate (but conditional) predictive means.

In the application of person-pair imputations, none of the variables imputed were part of a multivariate set, so that it was not necessary to set up a hierarchy of variables for the series of conditional models described above. Instead, each pair variable was imputed one at a time, where the only multivariate application was the necessity to have a multivariate predicted mean vector when the response was polytomous. What is given below is an abbreviate description of the method in the univariate case only. A description in the multivariate case is given in the NSDUH imputation report (Grau et al., 2005).

N.3 Outline and Description of Method

The procedure for implementing PMN in the 2003 NSDUH, where imputed variables were not part of any multivariate set, entailed five steps, which are listed below:

N.3.1 Step 1: Setup for Model Building and Hot-Deck Assignment

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

N.3.2 Step 2: Modeling

The model was built using the complete data respondents only with weights adjusted for item nonresponse.

N.3.4 Step 3: Computation of Predictive Means and Delta Neighborhoods

Once the model was fitted, the predictive means for item respondents and item nonrespondents were calculated using the model coefficients. This predictive mean (or predictive mean vector, in the polytomous response case) was the matching variable in a random NNHD.

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

The pool of donors was further restricted to satisfy constraints to make imputed values consistent with the preexisting nonmissing values of the item nonrespondent. An example of this type of constraint, called a "logical constraint," was given by the pair relationship in the imputation of multiplicities. Other constraints, called "likeness constraints," were placed on the pool of donors to make the attributes of the neighborhood as close to that of the recipient as possible. For example, for the imputation of pair relationships, donors and recipients among pairs where both respondents were in the 21-to-25-year-old age range were restricted to have had the same or similar marital status whenever possible. A small value of delta could have also been considered as a likeness constraint. Whenever insufficient donors were available to meet the likeness constraints, including the preset small value of delta, the constraints were loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors were available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor was found using a sequential hot deck, where matching was done on the predictive mean. (Even though weights would not have been used to determine the donor in the sequential hot deck, "unweighted" is not an accurate characterization of the imputation process because weighting would already have been incorporated in the calculation of the predicted mean.)

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

N.3.4 Step 4: Assignment of Imputed Values Using a Univariate Predictive Mean Neighborhood

Using a simple random draw from the neighborhood developed in Step 4, a donor was chosen for each item nonrespondent. If only one response variable was imputed, the assignment step was a simple replacement of a missing value by the value of the donor.

With the MPMN method, the neighborhood was defined based on a vector of predictive means rather than from a single predictive mean as in the univariate case. In the case of the pair analyses, this vector encompassed a vector of predictive means from a single categorical model. For each item nonrespondent, a distance was calculated between the elements of this vector of predictive means, where the observed values were missing, and the corresponding elements of the vector for every item respondent. A neighborhood that resulted from this vector of distances was constrained by a multivariate preset delta, where the distances associated with each element of the predictive mean vector must each have been less than the preset delta associated with that element. From the donors that remained, a single neighborhood was created out of a vector of differences by converting that vector to a scalar, called the Mahalanobis distance, which is given by

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

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

As stated in the previous section, logical constraints were placed on the multivariate neighborhood, so that imputed values were consistent with preexisting nonmissing values. The k smallest Mahalanobis distances who met the logical constraints among all candidate donors for a given item nonrespondent were selected for the neighborhood. In addition to the multivariate delta, likeness constraints were used to make the donors in the neighborhood as much like the recipient as possible. These could have been loosened if insufficient donors were available.

As with the univariate assignments, a donor was randomly drawn from the neighborhood for each item nonrespondent. For most variables, the observed value of interest was donated directly to the recipient.

N.4 Comparison of PMN with Other Available Imputation Methods

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

- **Ability to use covariates to determine donors is far greater than in the hot deck.** As with other model-based techniques, using models allows more covariates to be incorporated, including measures of use of other drugs, in a systematic fashion, where weights can be incorporated without difficulty. However, like a hot deck, covariates not explicitly modeled can be used to restrict the set of donors using logical constraints. If there is particular interest in having donors and recipients with similar values of certain covariates, they can be used to restrict the set of donors using likeness constraints even if they are already in the model.
- **Relative importance of covariates is determined by standard estimating equation techniques.** In other words, there are objective criteria based on methodology, such as regression, that quantify the relationship between a given covariate and the response variable, in the presence of other covariates. Thus, the response variable itself is indirectly used to determine donors.
- **Problem of sparse neighborhoods is considerably reduced, which makes it easier to implement restrictions on the donor set.** Because the distance function is defined as a continuous function of the predictive mean, it is possible to find donors arbitrarily close to the recipient. Thus, it is less likely to have the problem of sparse neighborhoods for hot decking. Moreover, having sufficient donors in the neighborhood allows for imposing extra constraints on the donor set, which would be difficult to incorporate directly in the model.
- **Sampling weights are easily incorporated in the models.** The weighted hot deck can be viewed as a special case of PMN.
- **Correlations across response variables are justified by making the imputation multivariate.**
- **Choice of donor can be made random by choosing delta large enough such that the neighborhood is of a size greater than 1.** Under the assumption that the recipient and the candidate donors in the neighborhood have approximately equal means, the random selection allows the case where the error distribution with mean

zero can be mimicked. This helps to avoid bias in estimating means and totals, variances of which can be estimated as in two-phase sampling or by suitable resampling methods.

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

As noted earlier, the PMN method has some similarity with the predictive mean matching method of Rubin (1986) except that, for the donor records, the observed variable value and not the predictive mean is used for computing the distance function. Also, the well-known method of nearest neighbor imputation is similar to PMN, except that the distance function is in terms of the original predictor variables and would often require arbitrary scaling of discrete variables. Moreover, for this method, it is generally hard to objectively decide about the relative weights for different predictor variables.

Appendix O: Rules for Determining Pair Relationships

Appendix O: Rules for Determining Pair Relationships

O.1 Rules for determining matching pairs, in priority order

The following rules are used to determine the roster member in a respondent's household roster that corresponds to the other pair member. In the rules that follow, 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. In the following table, if the rules for Pair Member A and Pair Member B in a single row differ, 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. The rules, called priority conditions because of their hierarchical nature, are listed in priority order in Exhibit O.1, along with the number of pairs to which each rule was applied. Since the 2001 survey, 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. Rules involving situations where more than one MBRSEL existed are therefore not included in this table. Some other conditions which were not manifest in 2003 are also excluded from this table, provided the distribution of counts would have been unaffected by their exclusion from the code.

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
1	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	17,069
2	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	1,953
3	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	117
4	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	223
5	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	36
6	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	2

(continued)

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
7	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	252
8	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	21
9	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	2
10	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	13
11	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	21
12	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
13	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	2
14	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	34
15	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	3
16	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	1
17	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	1
18	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	173

(continued)

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
19	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	25
20	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	4
21	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	2
22	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	3
23	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	13
24	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	1
25	Age and gender match exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	3
26	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	1
27	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	3
28	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	1
29	Age within one, gender off, one MBRSEL	Age within one, gender off, one MBRSEL, only two in household	1

(continued)

Exhibit O.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
30	No match, but no relationship codes are missing, and none involve domains of interest	No match, but no relationship codes are missing, and none involve domains of interest	12
31	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	18
32	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	10
33	No match at all	No match at all	8

O.2 Rules for identifying pair relationships among pairs

Table O.2 summarizes the rules used to identify the pair relationships, using the relationship codes and questionnaire ages of the two pair members. Because the child (12 to 17)-parent and child (12 to 20)-parent relationships can be derived from relationships created using 12 to 14 year olds, 15 to 17 year olds, and 18 to 20 year olds, these latter relationships are the ones referenced in the rules. The variable PAIRREL, which is the next to last column of the table, identifies the pair relationship as defined by Table 6.1 in the main body of this report. As with the rules for identifying which members of the roster belong to the pair, these rules, also called priority conditions because of their hierarchical nature, are given in priority order. In the headers, the moniker "A" refers to pair member A, and "B" refers to pair member B. The relationship between A and B is described in the columns "A-B Relationship," from the perspective of pair member A ("B to A, according to A") and the perspective of pair member B ("A to B, according to B"). Any constraints on the pair members (other than FIPE3) are given in the columns "Constraint on A" and "Constraint on B." These constraints include age constraints, where a range of ages (e.g., 12-17) indicates that the value of the questionnaire edited age (AGE) is between the numbers given. Also in this column, "child" and "children" are defined as (a) roster member(s) with nonmissing ages smaller than 18. The question FIPE3 asks if the respondent is the parent of a selected 12 to 17 year old. The responses given in the table are either "yes" or "no." The column for RELMATCH indicates the quality of the match between pair members, as defined in Table 6.4 in the main body of this report. In the table, blank cells mean that no restrictions were placed on that variable to determine the pair relationship.

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
1	parent	child	12-14				1	1
	child	parent		12-14				
2	parent	child	15-17				2	1
	child	parent		15-17				
3	parent	child	18-20				3	1
	child	parent		18-20				
4	parent	child	21+				4	1
	child	parent		21+				
5	sibling	sibling	12-14	15-17			5	1
	sibling	sibling	15-17	12-14				
6	sibling	sibling	12-17	18-25			6	1
	sibling	sibling	18-25	12-17				
7	sibling	sibling	no constraints, after considering #5 & #6				7	1
8	spouse/partner	spouse/partner	>=1 child	>=1 child			8	1
9	spouse/partner	spouse/partner	0 children, no bad data	0 children, no bad data			9	1
10	spouse/partner	spouse/partner	>=1 child	0 children, some bad data			8	1.5
	spouse/partner	spouse/partner	0 children, some bad data	>=1 child				
11	spouse/partner	roommate/nonrelative	>=1 child both sides, equal number each side				8	3
	roommate/nonrelative	spouse/partner	>=1 child both sides, equal number each side					

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
12	partner	partner	>=1 child	0 children, but other's children in household			8	3
	partner	partner	0 children, but other's children in household	>=1 child				
13	spouse/partner	spouse/partner	no constraints, after considering #8-#12				10	1
14	grandchild	grandparent					11	1
	grandparent	grandchild						
15	parent-in-law	child-in-law					12	1
	child-in-law	parent-in-law						
	other relative	other relative						
	roommate/boarder/ nonrelative	roommate/boarder/ nonrelative						
16	roommate/boarder/ other relative/ nonrelative/in-laws	roommate/boarder/ other relative/ nonrelative/in-laws					13	1
17	parent	missing	12-14	B less than 10 yrs. older th. A.		2	14	0
	missing	child						
18	parent	missing	12-14				1	2
	missing	child						
19	child	missing	A less than 10 yrs. older th. B.	12-14	2		14	0
	missing	parent						
20	child	missing		12-14			1	2
	missing	parent						
21	parent	missing	15-17	B less than 10 yrs. older th. A.		2	14	0
	missing	child						
22	parent	missing	15-17				2	2
	missing	child						

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
23	child	missing	A less than 10 yrs. older th. B.	15-17	2		14	0
	missing	parent						
24	child	missing		15-17			2	2
	missing	parent						
25	parent	missing	18-20				3	2
	missing	parent		18-20				
26	child	missing		18-20			3	2
	missing	child	18-20					
27	parent	missing	21+				4	2
	missing	parent		21+				
28	child	missing		21+			4	2
	missing	child	21+					
29	sibling	missing	12-14	15-17			5	2
			15-17	12-14				
	missing	sibling	12-14	15-17				
			15-17	12-14				
30	sibling	missing	12-17	18-25			6	2
			18-25	12-17				
	missing	sibling	12-17	18-25				
			18-25	12-17				
31	sibling	missing	no constraints, after considering #24, #25				7	2
	missing	sibling	no constraints, after considering #24, #25					
32	spouse/partner	missing	>=1 child	no spouse in roster			8	2
	missing	spouse/partner	no spouse in roster	>=1 child				

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
33	spouse/partner	missing	0 children, no bad data	no spouse in roster			9	2
	missing	spouse/partner	no spouse in roster	0 children, no bad data				
34	spouse/partner	missing	after #27, #28, no constraints	no spouse in roster			10	2
	missing	spouse/partner	no spouse in roster	after #27, #28, no constraints				
35	grandchild	missing	A at least 20 year older than B				11	2
	missing	grandparent						
	grandparent	missing	B at least 20 yrs older than A					
	missing	grandchild						
36	roommate/boarder/ other relative/ nonrelative/in-laws	missing			no		12	2
	missing	roommate/boarder/ other relative/ nonrelative/in-laws			no			
37	roommate/boarder/ other relative/ nonrelative/in-laws	missing					13	2
		roommate/boarder/ other relative/ nonrelative/in-laws						
38	nonmissing	child	12-14			yes	1	3
39	nonmissing	parent		12-14	yes		1	3
40	child	nonmissing		12-14	yes		1	3
41	parent	nonmissing	12-14			yes	1	3
42	nonmissing	child	15-17			yes	2	3

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
43	nonmissing	parent		15-17	yes		2	3
44	child	nonmissing		15-17	yes		2	3
45	parent	nonmissing	15-17			yes	2	3
46	parent	roommate/boarder/ othr relative/ nonrelative	12-14			no	13	3
						missing	15	4
	roommate/boarder/ other relative/ nonrelative	parent		12-14	no		13	3
					missing		15	4
47	parent	roommate/boarder/ other relative/ nonrelative	15-17			no	13	3
						missing	16	4
	roommate/boarder/ other relative/ nonrelative	parent		15-17	no		13	3
					missing		16	4
48	parent	roommate/boarder/ other relative/ nonrelative	18-20				17	4
	roommate/boarder/ other relative/ nonrelative	parent		18-20			17	4
49	parent	roommate/boarder/ other relative/ nonrelative	21+				18	4
	roommate/boarder/ other relative/ nonrelative	parent		21+			18	4

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
50	nonmissing not a sibling	child	12-14	21-75		no	13	3
			12-14, exactly one parent	21-75, exactly one spouse		missing	1	3
			12-14, 0 or 2 parents, or B has 0 or 2 spouse	21-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	15	4
	child	nonmissing not a sibling	21-75	12-14	no		13	3
			21-75, exactly one spouse	12-14, exactly one parent	missing		1	3
			21-75, 0 or 2 spouses, or A has 0 or 2 parents	12-14, 0 or 2 parents, or B has 0 or 2 spouse	missing		15	4
51	nonmissing not a sibling	child	15-17	24-75		no	13	3
			15-17, exactly one parent	24-75, exactly one spouse		missing	2	3
			15-17, 0 or 2 parents, or B has 0 or 2 spouse	24-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	16	4
	child	nonmissing not a sibling	24-75	15-17	no		13	3
			24-75, exactly one spouse	15-17, exactly one parent	missing		2	3
			24-75, 0 or 2 spouses, or A has 0 or 2 parents	15-17, 0 or 2 parents, or B has 0 or 2 spouse	missing		16	4

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
52	nonmissing not a sibling	child	18-20, exactly one parent	27-75, exactly one spouse		missing	3	3
			18-20, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	17	4
	child	nonmissing not a sibling	27-75, exactly one spouse	18-20, exactly one parent	missing		3	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	18-20, 0 or 2 parents, or B has 0 or 2 spouse	missing		17	3
53	nonmissing not a sibling	child	21+, exactly one parent	27-75, exactly one spouse		missing	4	4
			21+, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		missing	18	3
	child	nonmissing not a sibling	27-75, exactly one spouse	21+, exactly one parent	missing		4	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	21+, 0 or 2 parents, or B has 0 or 2 spouse	missing		18	4
54	spouse	sibling	one is 12-14, other, 15-17 both sides have parents or spouses				5	3
	sibling	spouse						
55	spouse	sibling	one is 12-17, other, 18-25 both sides have parents or spouses				6	3
	sibling	spouse						

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
56	spouse	sibling	ages neither 12-14/15-17 nor 12-17/18-25; both sides have parents or spouses				7	3
	sibling	spouse						
57	other relative	sibling	both sides have 2 parents; ages of oldest parents on either side differ by > 5 years; age of youngest parents on either side differ by > 5 years				13	3
	sibling	other relative						
58	nonmissing, not child	sibling	15-17	12-14			19	4
	sibling	nonmissing, not child	12-14	15-17				
59	nonmissing, not parent	sibling	12-14	15-17			19	4
	sibling	nonmissing, not parent	15-17	12-14				
60	nonmissing, not child	sibling	18-25	12-17			20	4
	sibling	nonmissing, not child	12-17	18-25				
61	nonmissing, not parent	sibling	12-17	18-25			20	4
	sibling	nonmissing, not parent	18-25	12-17				
62	nonmissing, not child	sibling	ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4
	sibling	nonmissing, not child	ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
63	nonmissing, not parent	sibling	ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21	4
	sibling	nonmissing, not parent	ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21	4

Exhibit O.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
64	sibling	roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	at least one is between 18 and 20				13	3
	roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	sibling	at least one is between 18 and 20					
65	sibling	unusual in-law code	12-20	26 or over			13	3
	unusual in-law code	sibling	26 or over	12-20				
66	spouse/ partner	not a child, parent, or sibling	>=1 child aged < 18	no spouse			22	4
	not a child, parent, or sibling	spouse/partner	no spouse	>=1 child aged < 18				
67	spouse/partner	not a child, parent, or sibling	15 or over, 0 children, no bad data	15 or over, no spouse			23	4
	not a child, parent, or sibling	spouse/partner	15 or over, no spouse	15 or over, 0 children, no bad data				
68	grandparent, grandchild	not grandparent, not grandchild					25	4
	not grandparent, not grandchild	grandparent, grandchild						
69	any codes	any codes	no constraints	no constraints			14	0

Appendix P: Priority Conditions for Creating Household-Consistent Covariates

Appendix P: Priority Conditions for Creating Household-Consistent Covariates

P.1 Household size

In Exhibit P.1, blank entries indicate that no conditions were required for that set of variables. The reported household size variable is QD54, and the edited household size variable is TOTPEOP, which cannot differ from the raw variable by more than 1. Any variable suffixed by "A" indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the suffix "B." For example, "QD54A" reflects the reported household size for pair member A. The quality-of-roster counts are considered in the column "any roster missing?" The variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages. The variables that appear in the table are TGOODAGA and TGOODAGB, the total number of cases 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 and over. Finally, the variable used to describe the screener household size is SHHSIZE. The conditions used to create the variable HHSIZE resulted in no missing values for this variable, so that no imputation was required. The first column in Exhibit P.1 gives the hierarchical priority condition, with the frequency of occurrence for each priority condition in parentheses.

Exhibit P.1 Priority Conditions Used to Create Household-Consistent Household Size

Priority Condition, Frequency	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
1 (19151)	equal, both > 1, both nonmissing	equal, both > 1, both nonmissing				TOTPEOPA
2 (0)	equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA<= QD54A	A: no		QD54A
3 (0)	equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGB<= QD54B	B: no		QD54B
4 (0)	equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGA = TGOODAGB TGOODAGA<= TOTPEOPA		SHHSIZE not equal to QD54A	TOTPEOPA
			TGOODAGA = TOTPEOPA		no condition	
5 (0)	equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA = TGOODAGB TGOODAGB<= TOTPEOPB		SHHSIZE not equal to QD54B	TOTPEOPB
			TGOODAGB = TOTPEOPB		no condition	
6 (0)	equal, both > 1, both nonmissing	within one of each other			SHHSIZE at least as large or larger th. screener roster, equal to QD54A[1]	SHHSIZE
7 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1, not equal to QD54B	QD54B >= TGOODAGB		SHHSIZE >= 2, closer to QD54B than TOTPEOPB	QD54B
8 (8)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB <= TOTPEOPB (no bad roster ages if equal)		SHHSIZE >= 2, TOTPEOPB is as close as QD54B	TOTPEOPB
9 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB <= SHHSIZE		TGOODAGB <= SHHSIZE	SHHSIZE
10 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1				TGOODAGB

Exhibit P.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition, Frequency	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
11 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1, not equal to QD54A B: missing or 1	QD54A >= TGOODAGB		SHHSIZE >= 2, closer to QD54A than TOTPEOPA	QD54A
12 (7)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGB <= TOTPEOPB (no bad roster ages if equal)		SHHSIZE >= 2, TOTPEOP (A) is as close as QD54A	TOTPEOPA
13 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGB <= SHHSIZE		TGOODAGB <= SHHSIZE	SHHSIZE
14 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1				TGOODAGB
15 (0)	both missing or 1	both missing or 1			SHHSIZE >= 2, SHHSIZE at least as large or larger th. screener roster	SHHSIZE
16 (13)	not equal, both > 1	TOTPEOP(B)=QD54 (B)	A: At least one age range variable less than min.[2] B: Age range variables all same or larger than min.			QD54B
		TOTPEOPA=QD54 (A)	B: At least one age range variable less than min. A: Age range variables all same or larger than min.			QD54A
17 (0)	not equal, both > 1		A: At least one age range variable less than min. B: At least one age range variable less than min..		Age range variables all same or larger than min.	SHHSIZE

Exhibit P.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition, Frequency	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
18 (728)	not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger th. screener roster, equal to QD54A	QD54A
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger th. screener roster, equal to QD54B	QD54B
19 (1)	not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger th. screener roster, equal to QD54A	A: Maxima for each age range between given count and min.
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger th. screener roster, equal to QD54B	B: Maxima for each age range between given count and min.
20 (0)	not equal, both > 1	not equal, both >1	TGOODAGA = TGOODAGB and TGOODAGA = QD54A	A: no B: no		QD54A
			TGOODAGA = TGOODAGB, and TGOODAGA = QD54B	A: no B: no		QD54B
21 (0)	not equal, both > 1	not equal, both >1	TGOODAGA = QD54A TGOODAGB > QD54B	A: no B: no	SHHSIZE > QD54B	QD54A
			TGOODAGB = QD54B TGOODAGA > QD54A	A: no B: no	SHHSIZE > QD54A	QD54B
22 (0)	not equal, both > 1	not equal, both > 1	TGOODAGA>GOODAGEA TGOODAGB>GOODAGEB TGOODAGA = SHHSIZE TGOODAGB = SHHSIZE	A: no B: no	TGOODAGA = SHHSIZE, TGOODAGB = SHHSIZE	SHHSIZE

Exhibit P.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition, Frequency	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
23 (0)	not equal, both > 1	not equal, both > 1	TGOODAGA > GOODAGEA TGOODAGB > GOODAGEB TGOODAGA = TGOODAGB	A: no B: no		TGOODAGA
24 (7)	not equal, both > 1	not equal, both > 1		A: no B: no	SHHSIZE=sum of maxima for each age group across pair members	SHHSIZE
25 (115)	not equal, both > 1	not equal, both > 1		A: no B: no	SHHSIZE >= 2, at least as large or larger than screener roster, closer to one of the QD54's	QD54A if SHHSIZE closer to A; QD54B if closer to B
					SHHSIZE >= 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member

Exhibit P.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition, Frequency	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
26 (1)	not equal, both > 1	not equal, both > 1		A fewer than B	SHHSIZE >= 2, at least as large or larger than screener roster, closer to QD54A than QD54B	QD54A
				B fewer than A	SHHSIZE >= 2, at least as large or larger than screener roster, closer to QD54B than QD54A	QD54B
				no condition	SHHSIZE >= 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member
				no condition	SHHSIZE >= 2, at least as large or larger than screener roster	SHHSIZE

[1]The abbreviation "th." stands for "than."

[2]The "min." refers to the minimum possible within each age range based upon the ages of the two pair members.

P.2. Age variables

Exhibit P.2 illustrates the hierarchical priority conditions ("priorities") used to create a new household-consistent 12 to 17 age group count; similar priority conditions are used for the 18 to 25, 26 to 34, 35 to 49, and 50+ age groups. In this table, blank entries indicate that no priority conditions were required for that set of variables. As with the previous set of tables, a variable followed by "A" (either in parentheses or not) indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the "B." As stated earlier, the variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages, and the variables TGOODAGA and TGOODAGB are also the total number of cases in the roster with valid ages, but if the original adjusted count is less than the minimum required, the original count is replaced by the minimum within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 and over. As noted in Section 6.2, these counts are adjusted so that the roster ages match what was entered in each pair member's questionnaire. Hence, AGE1217A is the adjusted count of 12 to 17 year olds for pair member A, and AGE1217B is the adjusted count of 12 to 17 year olds for pair member B. If AGE1217A or AGE1217B is less than the minimum possible, the count is replaced by the minimum, which is given by TAG1217A and TAG1217B respectively. Otherwise, AGE1217A and TAG1217A are equivalent, as are AGE1217B and TAG1217B. The sum of AGE011A, AGE1217A, AGE1825A, AGE2634A, AGE3549A, and AGE50PA is GOODAGEA. Similarly, the sum of AGE011A, TAG1217A, TAG1825A, TAG2634A, TAG3549A, and TAG50PA is TGOODAGA. The same can be said for GOODAGEB and TGOODAGB. The final 12 to 17 age count is denoted by AGE1217. The screener age count, denoted by SAGE1217, is only used if the age counts in each pair member's roster cannot conform to the minimum necessary, or are otherwise not possible to incorporate. If, after all edits, the count for AGE1217 is missing but the count for other age groups are not, and the counts for the 0 to 11 age group are the same for both pair members, then the sum of the counts for the other age groups, plus the minimum possible for AGE1217, are given by EXC1217. If other means fail to determine the appropriate value for the age count, match measures are used. These are measures which summarize the quality of the match between the two pair members. A match label of "0" indicates a perfect match, where the pair member's roster has a household member that is identified as the other pair member with a perfect match on age, gender, and is indicated as the other pair member by the MBRSEL variable. There are several levels of match measures, where a lower number signifies a better quality match. These measures are explained in detail in Section 6.2.2.1. As a final check, if the age group counts do not equal HHSIZE, and the count for the pair members are unequal, then the count is set to missing. As with Exhibit P.1, the first column in Exhibit P.2 gives the hierarchical "priority," with the frequency of occurrence for each priority in parentheses, for the AGE1217 count. In most cases, the frequencies corresponding to the other age ranges were the same as the frequency for AGE1217. In those cases where the frequency differed, footnotes give details of the differences.

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
1 (1)[1]	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE all nonmissing, all > 1	AGE1217A < min. (minimum), AGE1217B >= min.				AGE1217B
2 (0)[2]		AGE1217B < min. AGE1217A >= min.				AGE1217A
3 (0)		AGE1217A < min. AGE1217B < min.		SHHSIZE = HHSIZE SAGE1217 >= min.		SAGE1217
4 (6)[3]		AGE1217A = AGE1217B; both >= min.	Another count except 12-17 < min.			AGE1217A
5 (0)[4]		AGE1217A not equal to AGE1217B; both >= min.	AGE1825A < min.; AGE1825B >= min.			AGE1217 B
6 (0)[5]			AGE1825B < min.; AGE1825A >= min.			AGE1217A
7 (0)			Another count except 12-17 < min.		fewer roster entries missing in A than B	AGE1217A
8 (0)					fewer roster entries missing in B than A	AGE1217 B
9 (0)[6]					A & B: none missing A has better match measure than B	AGE1217A

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
10 (0) [7]	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE all nonmissing, all > 1	AGE1217A not equal to AGE1217B; both >= min.	Another count except 12-17 < min.		A & B: none missing B has better match measure than A	AGE1217B
11 (0)					A & B: none missing Age (A) >= Age (B)	AGE1217A
					A & B: none missing Age (B) > Age (A)	AGE1217B
12 (0)						missing
13 (18,472)		AGE1217A = AGE1217B	All other counts equal across pair members			AGE1217A
14 (438)		At least one age group has an unequal count between pair members		A: all age counts are equal to their screener counterparts	no missing roster entries on either side	AGE1217A
				B: all age counts are equal to their screener counterparts	no missing roster entries on either side	AGE1217B
15-22 (36)					A & B: none missing A has better match measure than B	AGE1217A
					A & B: none missing B has better match measure than A	AGE1217B
23 (76)					A & B: none missing Age (A) >= Age (B)	AGE1217A
					A & B: none missing Age (B) > Age (A)	AGE1217B

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
24 (0)	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE all nonmissing, all > 1				fewer roster entries missing in A than B A has good match measure (labels 0-7)	AGE1217A
					fewer roster entries missing in B than A B has good match measure (labels 0-7)	AGE1217B
25 (0)					fewer roster entries missing in A than B A is older than B	AGE1217A
					fewer roster entries missing in B than A B is older than A	AGE1217B
26 (0)					fewer roster entries missing in A than B B is older than A	AGE1217B
					fewer roster entries missing in B than A A is older than B	AGE1217A
27 (0)					A & B: same number of roster entries missing (>0) A is older than B	AGE1217A
28 (0)					A & B: same number of roster entries missing (>0) B is older than A	AGE1217B

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
29 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A < min., AGE1217B = min.				AGE1217B
		AGE1217B < min. AGE1217A = min.				AGE1217A
30 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 >= min.		SAGE-1217
31 (0)		AGE1217A = AGE1217B; both >= min.				AGE1217A
32 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B >= min.			AGE1217B
			AGE1825B < min. AGE1825A>= min.			AGE1217A
33 (0)					fewer roster entries missing in A than B	AGE1217A
					fewer roster entries missing in B than A	AGE1217B
34 (0)					A & B: same number of roster entries missing (>0) A has good match measure (labels 0-7)	AGE1217A
					A & B: same number of roster entries missing (>0) B has good match measure (labels 0-7)	AGE1217B

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
35 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A not equal to AGE1217B			A & B: same number of roster entries missing (>0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (>0) B is older than A	AGE1217B
36 (0)		priority conditions 29-35 not met				missing
37 (0)		AGE1217 missing after priority conditions 29-36 invoked, other age range counts not missing				HHSIZE – sum of other age counts
38 (354)		priority conditions 29-37 not met				AGE1217A
39 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A < min., AGE1217B = min.				AGE1217B
		AGE1217B < min. AGE1217A = min.				AGE1217A
40 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 >= minimum		SAGE-1217
41 (0)		AGE1217A = AGE1217B; both >= min.				AGE1217A
42 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B >= min.			AGE1217B
			AGE1825B < min. AGE1825A>= min.			AGE1217A
43 (0)					fewer roster entries missing in B than A	AGE1217B
					fewer roster entries missing in A than B	AGE1217A

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
44 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A not equal to AGE1217B			A & B: same number of roster entries missing (>0) B has good match measure (labels 0-7)	AGE1217B
					A & B: same number of roster entries missing (>0) A has good match measure (labels 0-7)	AGE1217A
45 (0)					A & B: same number of roster entries missing (>0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (>0) B is older than A	AGE1217B
46 (0)		priority conditions 39-45 not met				missing
47 (0)		AGE1217 missing after priority conditions 39-46 invoked, other age range counts not missing				HHSIZE – sum of other age counts
48 (483)		priority conditions 39-45 not met				AGE1217B

Exhibit P.2 Priority Conditions Used to Create Household-Consistent Age Variables (continued)

Priority Condition, Frequency	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
49 (78)	TGOODAGA = HHSIZE					TAG1217A
	TGOODAGB = HHSIZE					TAG1217B
50 (65)[8]	SHHSIZE equals HHSIZE	AGE1217A, AGE1217B <= SAGE1217		AGE1217A & B <= SAGE1217		SAGE1217
51 (1)[9]	SHHSIZE = HHSIZE; HHSIZE = EXC1217	AGE1217 missing	other counts not missing, AGE011A equals AGE011B			MIN1217
52 (10)[10]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, either the two TOTPEOPs > 0, or SHHSIZE=HHSIZE			AGE1217A equals SAGE1217		AGE1217A
				AGE1217B equals SAGE1217		AGE1217B

[1]The frequency of priority condition #1 for AGE1214, AGE1825, AGE2634, and AGE50p was 0. The frequency for AGE1220 and AGE3549 was 2.

[2]The frequency of priority condition #2 for AGE2634 was 2. The frequency for AGE3549 was 1.

[3]The frequency of priority condition #4 for AGE1214 was 7. The frequency for AGE1220 was 5. The frequency for AGE2634 and AGE3549 was 2.

[4]The frequency of priority condition #5 for AGE2634 was 2.

[5]The frequency of priority condition #6 for AGE3549 was 2.

[6]The frequency of priority condition #9 for AGE1825 and AGE50p was 1.

[7]The frequency of priority condition #9 for AGE2634 was 1.

[8]The following frequencies were observed for the other age range counts for priority condition #50: AGE1214, 66; AGE1220, 60; AGE1825, 63; AGE2634, 64; AGE3549, 68; AGE50p, 69

[9]The frequency of priority condition #51 for the age range counts other than AGE1217 was 0.

[10]The frequency of priority condition #52 for AGE1214 and AGE3549 was 9. The frequency for AGE2634 was 11. The frequency for AGE50p was 12.

Appendix Q: Multiplicity and Household Count Model Summaries

Appendix Q: Multiplicity and Household Count Model Summaries

Q.1 Introduction

The exhibits in this appendix list the covariates used in all the models that were run to impute missing values in the pair relationship, multiplicity, and household count variables. For each variable or set of variables to which the predictive mean neighborhood (PMN) imputation method was applied, three models were run: one to adjust the weights for item nonresponse (response propensity models), and a second and third to calculate predictive means. In the second model, household composition was represented by the household size variable, HHSIZE, and in the third, household composition was represented by the household composition age count variables. Imputation was sometimes performed within separate model groups, so that separate exhibits are required for those model groups.

Section Q.2 deals with the pair relationship variables; Section Q.3 deals with the multiplicity variables; and Section Q.4 deals with the household-level person count variables. In addition, when the initialism "MSA" is used, it represents "metropolitan statistical area." Finally, these models were at a pair level, whereas some of the variables in the models were at a person level. To differentiate which respondent the person-level variable applied to, the variable label is followed by a parenthetical "older" or "younger" to refer to the variable corresponding to the older or younger respondent, respectively. If the respondents in the pair were the same age, one of the respondents was randomly selected to be "older" or "younger."

Q.1.1 Screener and Segment-level Variables

In the PMN procedure, statistical modeling was performed to adjust weights for item nonresponse and also to calculate predictive means in the imputation models. Descriptions of questionnaire-derived variables are described in detail in the main body of the text. No such descriptions are available for screener and segment-level variables, however. The following screener and segment-level variables were often used as covariates in both types of models for the PMN procedures.

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.

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 the respondents lived in according to modified 1990 census data, which was adjusted to more recent data from Claritas, Inc.[20] This variable had five levels: segment in

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

metropolitan statistical area (MSA) with 1 million or more persons; segment in MSA with 250,000 to 999,999 persons; segment in MSA with fewer than 250,000 persons; segment not in MSA and not in rural area; and segment not in MSA and in rural area.

Percentage Hispanic Population

The percentage Hispanic population variable was used to categorize segments according to the concentration of Hispanics in the segments in which the respondents lived, using the adjusted 1990 census data. It had three levels: less than 20 percent, 20 to 70 percent, and more than 70 percent.

Percentage of Owner-Occupied Households

The percentage owner-occupied household variable was used to categorize segments according to the concentration of owner-occupied households in the segments in which the respondents lived, using the adjusted 1990 census data. It was used as a surrogate for income because wealthy segments tend to have many homeowners, while poor segments tend to have many renters. It had three levels: less than 10 percent, 10 to 50 percent, and 50 percent or more.

Percentage Black Population

The percentage black population variable was used to categorize segments according to the concentration of black or African American households in the segments in which the respondents lived, using the adjusted 1990 census data. It also had three levels: less than 10 percent, 10 to 40 percent, and 40 percent or more.

Q.2 Pair Relationship Variables

Exhibit Q.1 Model Summaries (Pair Relationships)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
0 (12-14, 12-14)	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Black in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
1 (12-14, 15-17)	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
2 (12-14, 18-25)	Household Size, Age Category (older), Race (older), Sex (younger), Employment (older), Census Region, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
3 (15-17, 15-17)	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
4 (15-17, 18-25)	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Marital Status (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
5 (18-20, 18-25)	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Age Category (older), Race (older), Sex (older), Marital Status (older),Marital Status (younger), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Age Category (older), Race (older), Marital Status (older), Education (older), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
6 (21-25, 21-25)	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Race (older), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
7 (12-14, 26+)	Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Employment (older), Census Region, MSA,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
8 (15-17, 26+	Household Size, Age Category (older), Race (older), Sex (older), Marital Status (older), Education (older), Employment (older), Census Region, MSA,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.1 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
9 (18-20, 26+)	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Black in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Education (younger) Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50p, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Marital Status (younger), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
10 (21+, 26+)	Age Category (older), Education (older), Education (younger), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment,	Household Size, Race (older), Sex (older), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Age Category (older), Race (older), Education (older), Education (younger), Employment (older), Employment (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Q.3 Multiplicities

Exhibit Q.2 Model Summaries (Multiplicities)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Parent-child (12-20) parent focus	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.2 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-14) Sibling (15-17) Younger Sibling Focus	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25) Older Sibling Focus	Household Size, Race (older), Sex (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25) Younger Sibling Focus	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Q.4 Household-Level Person Counts

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) child focus, both pair members < 18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus, at least one pair member older than 18	Household Size, Race (older), Sex (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) parent focus, both pair members <18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Parent-child (12-20) parent focus, at least one pair member older than 18	Household Size, Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Education (older), Marital Status (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, both pair members <18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, at least one pair member older than 18	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-17) Sibling (18-25), Older Sibling Focus, both pair members <18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment,	Household Size, Race (older), Sex (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, at least one pair member older than 18	Household Size, Race (older), Sex (older), Sex (younger), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse, both pair members < 18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Age Category (older), Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment

Exhibit Q.3 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Spouse-spouse, at least one pair member older than 18	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category (older), Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse with children, both pair members <18	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Number of Spouse-Spouse (what)	Household Size, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race (older), Sex (older), Sex (younger), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse with children, at least one pair member older than 18	Household Size, Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment	Race (older), Sex (older), Sex (younger), Marital Status (older), Education (older), Employment (older), Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-child (12-20) child focus, < 18	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) child focus, older than 18	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) parent focus, <18	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Race, Sex, Census Region, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Parent-child (12-20) parent focus, older than 18	Household Size, Age Category, Race, Sex, Education, Marital Status, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Number in Household Aged 0-11, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	**Not Including Household Size**
Sibling (12-14) Sibling (15-17), Older Sibling Focus, <18	Household Size, Race, Census Region, MSA, Categorical Percent Hispanic in Segment,	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, older than 18	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household, Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment	Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, <18	Household Size, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment,	Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment, Household Size	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Exhibit Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent is not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25), Older Sibling Focus, older than 18	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Household Size, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment
Spouse-spouse, < 18	Age Category, Race, Census Region,	Age Category, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment	Age Category, Race, Sex, Census Region, MSA, Categorical Percent Hispanic in Segment
Spouse-spouse, older than 18	Age Category, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Age Category, Race, Sex, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Age Category, Race, Sex, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment
Spouse-spouse with children	Household Size, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment,	Household Size, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment,	Number in Household Aged 0-11, Number in Household Aged 12-17, Number in Household Aged 18-25, Number in Household Aged 26-34, Number in Household Aged 35-49, Number in Household Aged 50+, Race, Sex, Marital Status, Education, Employment, Census Region, MSA, Categorical Percent Hispanic in Segment, Categorical Percent Black in Segment, Categorical Percent Owner-Occupied Households in Segment

Appendix R: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

Appendix R: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

R.1 Introduction

In order 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 is considered the direct count, and the count from the other pair member is considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided both rosters did not have bad data codes. The strategy was also usually clear if one pair member had bad data in the household roster; the count from the pair member with good data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was only selected if substituting the appropriate relationship codes for the bad data codes would have given 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. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to reconcile these disagreeing counts are outlined.

R.2 Parent-child counts

For parent-child counts, the screener and the FIPE3 variable were 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 reflected the selected child's parents, and were limited to have values of 1 or 2. If neither side had bad relationship codes and the direct count was 2 while the indirect count was 1, the following rules applied:

1. The direct count might have exceeded the indirect count because one parent had left or entered the household between interviews. In this case, the ages in the rosters were matched to the screener roster to determine which count to believe.

2. The direct count might have exceeded the indirect count because the selected parent did not consider the other "parent" a spouse or live-in partner. If the pair relationship was not imputed, the indirect count was selected. However, if the pair relationship was imputed and the older pair member called the younger pair member a child, then the older pair member considered the child's "true" parent as not a spouse or live-in partner, even though he/she claimed the "true" parent's children. In this case, the direct count was used (the child's adjusted count).

If the direct count was 1 but the indirect count was 2, the child only listed one parent, but the parent lists a spouse (a "stepparent") or live-in partner in the household roster. The following rules applied:

1. The indirect count might exceed the direct count because the selected child did not accept a stepparent or live-in partner as his/her parent. If this stepparent or live-in partner was the other respondent selected, we determined this was a child-parent pair based on the response of the "parent" to the FIPE3 question. If the FIPE3 question was answered "yes," the RELMATCH variable had a value of 3, and the indirect count was selected as the multiplicity count. If the FIPE3 question was answered "no," the pair was not considered a child-parent pair, and would not be considered for these counts. Finally, if the FIPE3 question was not answered, the respondent was considered a "parent" if he or she was a stepparent. If the respondent was a live-in partner, the determination of the pair relationship was left to imputation. The multiplicity count was set to the indirect count to account for the possibility that the pair relationship would be imputed as parent-child.

2. Suppose the selected child did not accept a stepparent or live-in partner as his/her parent (as above), but the other respondent selected was the "true" or "original" parent. In this case, the stepparent or live-in partner was only identified in the "original" parent's roster, so there was no way to determine how the stepparent or live-in partner would have answered the FIPE3 question. The stepparent was considered a "parent" even if the child did not view him or her this way, so that the indirect count was used. The case of live-in partners was less clear. If the live-in partner had been selected, the determination of whether a parent-child relationship was indicated would have involved the response to the FIPE3 question, which we didn't have since the live-in partner was not selected. Hence, these cases were left to imputation.

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. If neither side had bad relationship codes, the following rules applied:

1. In most cases, if one pair member had bad data, the multiplicity was obtained from the other pair member. The exception was when the number of household members between 12 and 14, 12 and 17, 12 and 20, or 15 and 17, (depending on the domain) in the "bad side" matched the number in the corresponding age ranges in the screener roster, but the "good side" had a larger number in the corresponding age ranges than in the screener roster, a larger number which matched the multiplicity count of the "good side." The larger number was due to the fact that the "good side" originally had no "self" identified in its roster, and an extra roster member was incorrectly added to the "good side" in the roster editing stage to create a "self" (see Section 6.2.2.2 in the main body of the report). In this case, the count of individuals within the age range on the "bad side" was used as the final count. (This was only an issue in the 1999 survey year.)

2. If the count of children in the household within the relevant age ranges differed between the pair members, but one side had a count of children equal to the same count from the screener roster, the multiplicity count that corresponded to the pair member with the same count of children as the screener was used.

3. If the count of children in the household within the relevant age ranges differed between the pair members, and both sides had a multiplicity count that exceeded the count of all children from the screener roster, the number of children in the screener roster was used as the multiplicity count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

4. The direct count and indirect count might differ because either the child lists a sibling that the parent considers "another relative," or the parent lists a child that the child considers "another relative." In either case, since the parent was the one to answer the FIPE3 question, the multiplicity count from the parent's perspective was selected as the final count.

R.3 Sibling-sibling counts

Although there were two types of sibling-sibling pairs under consideration, each associated with two domains, the same rules could be 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). The younger age ranges in these rules can be switched to the older age ranges when the younger sibling was the focus. The following general rules apply:

1. The counts disagreed if a household member left or entered the household between interviews. As before, the roster that was closest to the screener was used to determine the count. If one roster member had the same number of household members within the ages of 12 to 14 or 12 to 17 (depending on the domain) as the screener roster, the multiplicity count from that roster member was used, provided the member had no bad relationship codes within the relevant age range.

2. If the counts disagreed and both exceeded the screener count of household members within the relevant age range, the multiplicity count was set to the screener count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

3. If the younger pair member identified the older as "sibling" but the older pair member did not reciprocate, then imputation was required to establish whether the relationship was sibling-sibling. For those pairs that were imputed to sibling-sibling, the count was incremented by 1 to reflect the fact that the younger sibling's relationship code was changed from nonsibling to sibling. However, if the younger sibling identifies other siblings within the relevant age range that the older sibling did not identify, then it was necessary to accept the direction of the imputation–that is, to identify these other roster members as siblings.

4. The counts disagreed if the siblings disagreed on whether one or more household members within the relevant age range was a sibling of theirs. However, if the minimum number of respondent's children possible, considering age ranges and sibling codes within both questionnaire rosters and the screener, was equal to the

maximum number possible, then the counts were set to the equal bounds. Otherwise, there was no way to reconcile these differing counts, so the final count was left to imputation, within the bounds determined by the two pair members' counts.

5. Other counts that were left to imputation involved cases where both sides had too many bad relationship codes to definitively determine a multiplicity count.

Appendix S: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

Appendix S: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

S.1 Introduction

Household-level person counts for a particular domain were obtainable using the multiplicity counts if the pair belonged to a pair relationship that fit into that domain, provided only one family unit was in the household. No reconciliation between pair members was necessary in that case, since the reconciliation had already been done with the multiplicity counts. Other counts were obtained from single respondents, for whom no reconciliation was necessary. This appendix discusses the conditions used to reconcile differing household-level person counts when the pair belonged to a pair relationship that corresponded to different pair domains than the one being counted. Typically, the counts between the two pair members were in agreement, and the determination of the final household-level count was straightforward, provided both rosters did not have bad data codes.[21] The strategy was also usually clear if one pair member had bad data in the household roster; the count from the pair member with good data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was only selected if substituting the appropriate relationship codes for the bad data codes would have given 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. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to reconcile these disagreeing counts are outlined. For each pair domain, a set of general rules are given, each with specific conditions required for the general rule to be implemented. Within each general condition, if at least one of the specific conditions was not satisfied, upper and lower bounds were determined and the final count was left to imputation.

S.2 Parent-child counts

For parent-child counts where the pairs were not parent-child pairs of interest (e.g., sibling-sibling pairs, parent-child pairs where the child was 21 or over, etc.), the screener was used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. For the child-focus counts, the count is of the number of children of a parent in the household. The following general rules applied:

1. Among non-parent-child pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides:

---

[21] If a roster pointed to a household size of 1, this was considered "bad data," since both pair members in the household were survey respondents.

- Either no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or the counts were equal to the screener age counts, or a side with good data indicated siblings within the relevant age range living together in a household without parents.
- No situations where parents were not identified in the household, but some in the household had bad relationship codes and were old enough to be parents.
- No counts of one child in the relevant child-age range when both members of the pair were in that range, and the children were siblings.
- No pairs where the ages of the identified parents did not match and both sides had relationship codes indicating "other relative" or a nonrelative, indicating more than one family unit in the household.[22]
- The household size was greater than 1 and non-missing on both sides.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

   - If the number of children matched across both rosters and the screener.
   - If the counts which agreed with each other equaled or exceeded the count of the number of children from the screener.
   - If both sides had a count of 0, both had a roster, and (at least) one side had all good age and relationship codes.
   - If both sides had a count of 0, both had a roster, and the number of respondents who were old enough to be parents in the household was 0 according to the screener.

3. The counts might have agreed with a value of 1. If both pair members were children within the relevant age range, and both indicated they had parents even though the children were siblings, then they were not included in each other's rosters, but were obviously in the screener roster, so the count was set to 2.

4. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

   - Either:
     - There were no bad relationship codes within the relevant child-age ranges and the respondent identified parents in the household, or
     - There were no children within the relevant age range, or
     - No parents were identified in the household and nobody in the roster older than the respondent had a bad relationship code.

[22] Codes which indicate "other relative" or a non-relative are 7 (roommate), 8 (child-in-law), 10 (parent-in-law), 12 (boarder), 13 (other relative), and 14 (other non-relative).

- No counts of one child were in the relevant child-age range when both members of the pair were in that range, and the children were siblings.

5. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count, if that count was zero, under any of the following conditions. Either

   - The other roster was valid, did not have any bad ages, and had no ages in the relevant age range, or
   - The other roster was also bad, but the screener roster was valid, and did not have any ages in the relevant age range, or
   - The respondent identified both grandchildren and grandparents in the roster, where the grandchildren referred to the grandparent instead of the respondent.

6. When two different family units were in the household, the determination of the final count had to be treated separately. This could have included the multi-generational families referred to earlier, and to two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one might be zero) was used provided the following conditions were satisfied on both sides:

   - There were no bad ages or relationship codes within the relevant age ranges.
   - Both had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side.
   - The number of identified parents were not equal to the total number over 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes.
   - The number of identified children were not equal to the total number within the relevant age range in the household on either side, meaning that children with parents could correspond to roster members identified by other relationship codes.
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

If the above conditions were not met, the two families in the household might have been already accounted for when the counts were determined for each side. The maximum of the two counts was taken if the household members in the roster over 25 (of parental age) were either both equal to the number over 25 in the screener roster, or both different than the number over 25 in the screener roster. However, if the number over 25 in the screener roster was equal to the number over 25 in one of the pair member's rosters, but not the other, then the count where the number was equal to the screener roster was taken. In both instances, the count corresponding to the pair member that matched the screener roster was taken.

7. If one pair member did not have a valid roster, and the pair member with a valid roster was within the valid age range and was a sibling to the other pair member, but the count from his roster was only 1, the count was set to 2.

8. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts and the other did not, it was necessary to decide who to believe. Often this occurred because one of the respondents was a relative outside the nuclear family unit, like a cousin or aunt/uncle, whose own parents did not live in the household, or a boarder.[23] Selecting either the zero count or nonzero count in this instance required that:

   - The respondent with zero count did not identify parents in the roster or he/she identifies parents but was over 20 years old, and had no bad relationship codes within the relevant age ranges.
   - Either the respondent with nonzero count had siblings or children within the relevant age range, or was himself/herself within that age range (with a count of 1).

   When one count was zero and the other nonzero, the nonzero count was used under the following conditions

   - The respondent pair member with nonzero count did not have bad relationship codes,
   - Either:
     - The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, or
     - The count of children in the household within the relevant age range for the nonzero count pair member matched that of the screener, or
     - The count of children in the household within the relevant age range for the zero count pair member matched that of the screener, because a child was (or children were) listed as 11 years old in the nonzero count pair member's roster, when he or she (they) should have been 12 (according to the nonzero count pair member's and the screener roster) so that the final count was the nonzero count with this child (these children) added, or
     - The respondent with zero count had no household members with a family-type relationship code, or
     - The count of children within the relevant age range in the household for the zero count was closer to the screener age count, but the nonzero count, was less or equal to than the screener age count, or
     - The other conditions had not already established a nonzero count, but a count for a subset age group had already been established as nonzero. For example,

[23] Even if there was disagreement between the respondents about whether a boarder or other family member was in fact a sibling, parent, or child, this would had been resolved at the pair relationship stage, where we would had determined whether this was in a domain of interest.

if the count for 12 to 14 year olds was nonzero, then the 12 to 17 year old count had to be nonzero.

The zero count was used if:

- The household age composition among the relevant age ranges for the zero count pair member more closely matched the screener, or
- The pair was a grandparent-grandchild pair with an adult child of the grandparent living in the household. The nonzero count resulted from an assumption that a respondent's adult child and grandchild within the relevant age range were a parent-child pair. If the grandchild identified the grandparent's child as "other relative," and did not identify any parents, this indicated that the grandparent's adult child was an uncle/aunt of the grandchild, not a parent.

9. Even with sibling-sibling and parent-child pairs, sometimes one side had a zero count and the other had a nonzero count. This was usually due to one pair member having missing relationship codes for the roster member that would have been identified as a parent (i.e., relationship codes for roster members in a parental age range). If the count for the pair member with the entirely good roster was equal to the number within the appropriate age range for the pair member with bad relationship codes in the roster, the nonzero count was selected.

10. The two counts might have disagreed because one side had bad relationship codes within the relevant age range, and the other did not. If the sum of the number of bad relationship codes with the smaller count equaled the larger count, the larger count was chosen.

11. The two counts might have disagreed because they disagreed on the ages of one or more household member, even though each respondent's count included all the children in their respective roster. If the roster for one respondent more closely matched the screener in terms of the distribution of ages within the roster, then that respondent's count was chosen. If the screener roster was a valid roster, but had fewer children in the relevant age range than the nonzero count of either pair member, then the final count was set to the number of children in the relevant age range of the screener roster.

12. The two counts might have disagreed because they disagreed on the ages of one or more household member, and each respondent's count included all the children in their respective roster, but neither was closer to the screener count. If the screener count differed from each respondent's count by the same amount, was greater than one but less than the other, the screener count was used as the final count.

13. If the pair relationship was parent-child, and the parent-child counts were associated with the same age range, then the household-level person counts would have been obtained using the parent-focus multiplicity counts. However, this did not occur if the age range for the pair relationship differed from the age range for the parent-child counts. If the pair relationship was imputed to be parent-child, or it was

deemed parent-child even though the child did not consider the parent a "parent," but the parent answered the FIPE3 question, the nonzero count should be used as the final count.

14. If after all the above tests were done to find the final count, the minimum possible and maximum possible counts, considering both questionnaire rosters and the screener roster, were the same, then the final count was set to that value.

15. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

Parent-child pairs, parent focus. For the parent-focus counts, the count is of the number of parents of at least one child in the household The child-focus parent-child counts are processed first, so if the child-focus parent-child counts are zero, it necessarily means that the parent-focus counts will also be zero. Nonzero child-focus counts also point to nonzero parent-focus counts. After setting counts to 0 where necessary, the following general rules applied:

1. Among non-parent-child pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides.

   - No situations where both pair members were children in the relevant age range, but were in a spouse-spouse pair relationship, and both identified the same roster member as parent,
   - Either:
     1. No bad relationship codes for household members of an age to be parents, or
     2. The total count was 2, for 2 parents, or
     3. The total count + the number of grandparents equaled the total number 26 or over in the household, according to the screener roster.
   - The household size was greater than 1 and non-missing on both sides.

   Note that it was not necessary to check for bad relationship codes in the child age ranges, since it was already known that the count had to be at least 1, and the number of children was not important for the parent counts.

2. The counts may have agreed even though the above conditions were not met. The final count could still have been set to one of the sides if it was a sibling-sibling pair, and the bad codes in the parental age range were on one side only. This would indicate that the side with bad codes were not missing parental codes.

3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count if there were no bad relationship codes and no roster members with bad age and bad gender values. Other circumstances called for setting the final count to zero, which would necessarily be the case if the child-focus counts were zero.

4. When two different family units were in the household, the determination of the final count had to be treated separately. This could have included multi-generational families, or two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one might be zero) was used under the following conditions:

   - There were no bad ages or relationship codes within the relevant age ranges.
   - Both had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side.
   - The number of identified parents were not equal to the total number over 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes.
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

   If the above conditions were not met, the two families in the household might have been already accounted for when the counts were determined for each side. The maximum of the two counts was taken if the household members in the roster over 25 (of parental age) were either both equal to the number over 25 in the screener roster, or both different than the number over 25 in the screener roster. However, if the number over 25 in the screener roster was equal to the number over 25 in one of the pair member's rosters, but not the other, then the count where the number was equal to the screener roster was taken. In both instances, the count corresponding to the pair member that matched the screener roster was taken.

5. If the pair relationship was a spouse-spouse pair, and one of the pair members had a positive count, with an age within the relevant child age-range, then the count for that pair member was taken as the final count, provided there were no bad relationship codes in that roster for roster members aged 18 or over.[24]

6. The two counts might have disagreed with one count nonzero, and the other equal to 0. In order to make it to these conditions, the count had to be nonzero. The nonzero was chosen as the final count if:

   - The count was 1, and there were no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or
   - The count was 2.

[24] In almost all cases (all cases in the 2002 survey year), either the count for the other pair member was 0, or the count for the pair members was equal. In the latter case, one of the identified parents should have been "parent-in-law." In one case in the 2001 survey year, the counts were both nonzero and unequal. In that instance, this condition should not have been invoked. The software has been corrected for the 2003 survey year.

7. The two counts might have otherwise disagreed, where the number of roster members 26 or over disagreed between the two pair members. In these situations, one count was 1, and the other 2. The final count corresponded to the pair member with the number of roster members 26 or over closest to the screener number of roster members 26 or over, under the following conditions:

   - The difference between the screener count of the number of household members 26 or over, and the pair members' counts of this number of household members was not the same between the two pair members.
   - Neither pair member had bad ages in their rosters.
   - Each pair member either had no bad relationship codes in his or her roster, or had a nonzero count with no bad relationship codes among respondents 26 or over.

8. The two counts might have otherwise disagreed if the bad relationship codes referred to missing parental codes. If one side had no bad relationship codes, then the sum of the number of bad relationship codes and the count on the side with the bad codes was equal to the count on the side with no bad relationship codes.

9. The two counts might have disagreed where one count was 2, and the other was 3. Since households with two family units had already been considered, the maximum number of parents possible was 2, so the final count was set to 2.

10. If after all the above tests were done to find the final count, the minimum possible and maximum possible counts, considering both questionnaire rosters and the screener roster, were the same, then the final count was set to that value.

11. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

S.3 Sibling-sibling counts

The logic for the sibling-sibling counts did not depend upon whether the younger age range was 12 to 14 or 12 to 17, or whether the older age range was 15 to 17 or 18 to 25. It also did not depend upon which pair member was the focus, though for the household-level person counts, the older member focus counts were the only ones considered. Hence, the counts that are of interest are of roster members in the older age range. As with the parent-child pairs, the multiplicity counts could be used if the pair relationship was a sibling-sibling pair of interest. However, the counts had to be determined for all other pairs. The rules follow below, separated by the member of focus:

1. Among pairs that were not sibling-sibling pairs of interest, in most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides:

- The pair could not be a sibling-sibling pair, where both respondents were in the older age range, and have a younger sibling in the younger age range, and the count was 1. (This refers to a sibling-sibling pair that would not constitute a domain of interest.)
- No bad relationship codes in the lower range if the count was 0.
- Either:
  - No bad relationship codes in the upper range, or
  - The count matched the screener age count.
- The household size was greater than 1 and nonmissing on both sides.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

   - If the number of children matched across both rosters and the screener, for both the upper and lower age ranges.
   - If the count was 0, and one of the two was true:
     - Neither side had bad relationship codes or ages, or
     - The number of household members in the screener 26 years of age or older was 0.

3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

   - No bad relationship codes within the lower age range when the count was zero.
   - Either:
     - There were no bad relationship codes within the upper age range, or
     - The count was equal to the screener age count within the upper age range, or
     - The count was zero, and the count of household members in the lower age range was zero.

4. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count, under the following conditions:

   - The count was zero,
   - Either:
     - The number of children in either the lower or upper age ranges was zero with no bad ages in the roster, or
     - The number of children in the screener roster in either the lower or upper age ranges was zero, with a valid screener roster.

5. When two different sets of siblings were in the household, the determination of the final count had to be treated separately. The two sets of siblings refer to siblings where both parents from one set differ from the parents of the other set. The sum of the two counts (one might be zero) was used, provided the following conditions were satisfied for both pair members:

    - There were no bad relationship codes within the upper age ranges.
    - There were no bad relationship codes within the lower age range, or the count was nonzero.

6. If the counts from the two pair members did not agree, the following rules were used to assign the appropriate count, provided no bad relationship codes were evident in either age range, on either side. These conditions are hierarchical, in that subsequent conditions require that the previous condition was not met.

    - If the number within the upper age range was the same on both sides, but the number in the lower age range was not, the side was chosen with the number in the lower age range equal to the number in the screener roster within the lower age range. (In all cases, one size had zero count and the other did not. This captured situations where it was necessary to discern whether the zero count was due to no children in the lower age range on one side, and whether the screener also had no children in that range.)
    - For one pair member, the number of children in either the lower age range or the upper age range did not agree with the number in the screener in that range. However, for the other pair member, the number within both age ranges agreed with the screener count. The count was set to the side that agreed with the screener.
    - For both pair members, the numbers within the lower age range were either both zero, or both positive. The number within the upper age range did not agree between pair members, but one pair member agreed with the screener. The count was set to the count for that pair member.
    - In the rosters for both pair members and the screener, the numbers within the upper age range nonzero for at least one of the three were nonzero, but not necessarily equal. The numbers within the lower age range were not equal across any of the three rosters. The pair member with the number of children in the younger age range closest to the screener was selected.
    - In the rosters for both pair members and the screener, the numbers within the lower age range nonzero for at least one of the three were nonzero, but not necessarily equal. The numbers within the upper age range were not equal across any of the three rosters. The pair member with the number of children in the upper age range closest to the screener was selected.

7. If the counts from the two pair members did not agree, but one side had bad relationship codes within the upper age range, and the other did not have bad relationship codes, and the sum of the count and the number of bad relationship codes

on one side was equal to the count for the pair member with the good roster, the count for the pair member with the good roster was selected.

8. If the counts from the two pair members did not agree, but the above conditions were not met, in many cases this was due to one of the pair members not being part of the immediate family unit, in which case his or her count was automatically zero. To identify these cases, and assign the count to the other pair member, the following conditions had to be satisfied:

   - The pair relationship did not indicate an identifiable family-type relationship (e.g., sibling-sibling, parent-child, spouse-spouse, or grandparent-grandchild relationship).
   - Either:
     - One pair member did not have any relationship codes indicating parent, child, sibling, spouse, grandchild, or grandparent, and
     - The other pair member had at least one relationship code indicating a relationship other than parent, child, sibling, spouse, grandchild, or grandparent, and
     - For the pair member with family codes, either no bad relationship codes were within both the upper and lower age ranges, or no bad relationship codes were within the upper age range, and the count was positive,

     or

     - There were no bad relationship codes within both the upper and lower age ranges for either pair member.

9. If one pair member had no bad relationship codes within both the upper and lower age ranges, but the other had some bad codes, then the count associated with the pair member with no bad codes was selected if the count of immediate family members (parent, child, sibling, spouse, grandchild, grandparent) was the same as the count of household members within both the lower and upper age ranges.

10. If one pair member had a zero count due to having no household members within the upper age range, but the number of household members within that age range was nonzero for both the screener and the other pair member (though not necessarily equal), then a nonzero count was selected. If the count for the other pair member was equal to the number of household members within the upper age range for that pair member, then the final count was set to the screener number of household members within that age range.

11. If the pair was a spouse-spouse pair, one count might have been zero while the other was nonzero because the spouse-spouse pair still lived with the parents of one pair member, and the pair member's younger siblings also lived in the household. In this case, the nonzero count was selected if the number of immediate family

members (parent, child, sibling, spouse, grandchild, grandparent) if the roster for the pair member with the zero count was less than his or her total household size.

12. In some cases, one pair member called the other pair member a parent or child, but the other pair member did not reciprocate. In the case of a child who did not reciprocate the parent's identification of him or her as child, the child's count was always less than the parent's count. By the same token, in the case of a parent who did not reciprocate the child's identification of him or her as parent, the parent's count was always less than the child's count. If the pair relationship was imputed to be "parent-child," then the pair member who did not acknowledge a parent-child relationship was overruled, and the maximum count of the two pair members was selected as final.

S.3 Spouse-spouse counts (with or without children)

The multiplicity counts were not useful in the logic for the spouse-spouse household counts, since the spouse-spouse multiplicity counts were always 1. The logic for the spouse-spouse counts follows:

1. Among the majority of pairs, the counts for the two sides agreed. However, both sides had to meet the following conditions, in order for the final count to be set to one of the sides.

    - The pair could not be a spouse-spouse pair, where both respondents had a spouse or both respondents had a partner,
    - No bad relationship codes for roster members 15 or over,
    - The number of spouse-spouse pairs was either 1 or 0 for both pair members,
    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
    - The household size was greater than 1 and nonmissing on both sides.

    NOTE: This general condition failed to exclude some of the cases where one couple was identified by both pair members, but the identified couple was different for each pair member. This occurred most commonly with multgenerational families with two couples in the household, where the spouse/partner in the younger couple who "married into" the family did not recognize the spouse/partner's parents as parents-in-law. This has been corrected for processing in years subsequent to 2002.

2. The counts might have agreed even though the above conditions were not met. The count could still be set to one of the sides, if any one of the following was true:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.

- Either:
  - One pair member has a single bad relationship code, and no other relationship codes could match it to make it a couple (i.e., the pair member does not have a single identified parent, grandparent, parent-in-law, or child-in-law). The other pair member has no bad relationship codes.
  - One pair member has bad relationship codes among roster members 15 or over, or has bad ages, and the other has no bad ages or relationship codes, where the pair member with no bad roster entries has the same age composition as the screener. The pair member with the bad roster entries would have the same age composition as the screener if the number of roster members 15 or over was added to the number of roster members with bad ages.
  - One pair member has bad relationship codes among roster members 15 or over, or has bad ages, and the other has no bad ages or relationship codes, where all the relationship codes for the pair member with no bad roster entries are immediate family codes (child, parent, sibling, spouse, partner, grandparent, or grandchild). For the pair member with bad roster entries, all the existing relationship codes are immediate family codes.

3. If the household size was 1, or the number of respondents 15 or over in the household was 1 or 0, then the count should automatically be zero. Instead of setting the count to zero, the code set the count to pair member A's count. In a very small number of cases, this count was 1 instead of 0, which was an error.

4. For those cases where the pair was imputed to be a spouse-spouse pair, and both sides agreed that only one spouse-spouse pair was in the household, the count was set to one if:

   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
   - Either:
     - Both sides had fewer than 4 people older than 15 in the household, or
     - One side had fewer than 4 people older than 15 in the household, and the other had no bad relationship codes among roster members 15 or over.

5. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and
- Either:
  - There were no bad relationship codes among roster members 15 or over, or
  - There were no bad relationship codes among roster members 18 or over, and the pair member had parents.

6. When two different family units were already identified in the household, then two different parent-sets were being referenced (one of the parent-sets was often a single parent). The sum of the two counts (one might be zero) was used provided neither pair member had grandparents or grandchildren identified. This was to prevent spouse-spouse pairs from being counted twice, which would happen if grandparents were also parents of 0 to 17 year olds. If two family units were multigenerational families, then the final count was obtained by taking the maximum of the two pair members' counts.

7. It was possible for two different spouse-spouse pairs to be in the household, even though two different family units had not been identified. The final count was set to 2, even though two family units had not been previously identified, under the following conditions:

- The pair relationship was not a spouse-spouse pair, and the total household size was at least 4, and
- Either:
  - Both sides identified a spouse, or
  - Both sides identified a partner, or
  - One side identified a parent and the other identified a parent-in-law.

8. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could be because a couple entered the household or otherwise materialized after screening. The smaller count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
- The screener count of roster members 12 or over was no larger than the count of roster members 12 or over in the roster of the pair member with the smaller spouse-spouse count.

- The screener count of roster members 12 or over was smaller than the count of roster members 12 or over in the roster of the pair member with the larger spouse-spouse count.
- The difference between the screener count of roster members 12 or over and the count of roster members 12 or over in the questionnaire rosters of the pair members was smallest with the pair member with the smaller spouse-spouse count.

9. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could be because a couple left the household or otherwise dissolved after screening. The larger count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The screener count of roster members 12 or over was no larger than the count of roster members 12 or over in the roster of the pair member with the larger spouse-spouse count.
    - The screener count of roster members 12 or over was larger than the count of roster members 12 or over in the roster of the pair member with the smaller spouse-spouse count.

10. In many cases where the count of the number of spouse-spouse pairs did not agree between the two pair members, one side had zero count and the other did not. The nonzero count was selected if the pair member associated with the zero count was not a close relative, or somehow otherwise did not identify a spouse, partner, 2 parents, or 2 grandparents. The following conditions were required to select the nonzero count:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair member with a nonzero count either identified a spouse, a partner, two parents, or two grandparents.
    - The number of roster members 15 or over associated with the nonzero count pair member was no larger than the corresponding number associated with the zero count pair member.
    - If the side associated with the nonzero count identified a spouse, partner, or 2 parents, the following additional conditions were required:

- The number of roster members between 26 and 44 was the same between the two pair members.
- The number of roster members between 30 and 49 was the same between the two pair members.
- The number of roster members between 35 and 54 was the same between the two pair members.
- The number of roster members between 40 and 59 was the same between the two pair members.

- If the side associated with the nonzero count identified 2 grandparents, the following additional condition was required:
  - The number of roster members 50 or over was the same between the two pair members.

11. If either a pair member's partner was not considered a family member by the other pair member, or if a pair member had two grandparents, and an uncle/aunt husband-wife pair in the household, then the maximum was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - Either:
      1. At least one side identified a partner, and the maximum count was 1, or
      2. The pair member associated with the smaller count had a grandparent, and had at least 2 roster members who were not either a parents, siblings, children, spouses, partners, or grandparents.

    Note: this condition did not consider cases where the difference in counts was due to different household compositions between the pair members.

12. The count of the number of spouse-spouse pairs might not agree because one of the pairs was a sibling and sibling-in-law, and there are no codes for sibling-in-law. The maximum count was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair member with the smaller count did not have a spouse or partner, but did have siblings aged 15 or over, and there were household members in his or

her roster that were not parents, children, siblings, spouses, partners, grandchildren, or grandparents.

13. The count of the number of spouse-spouse pairs might not agree because one side had no nuclear family or grandparent-grandchild relationship codes, and one of the selected respondents was not in a child-parent, child-grandparent, or spouse-spouse relationship. The maximum count was selected if:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified,
    - The pair member's roster associated with the minimum count (usually zero) had no children, parents, siblings, spouses, partners, grandchildren, or grandparents among respondents 12 or over, and
    - The pair member's roster associated with the maximum count had some roster members who weren't children, parents, siblings, spouses, partners, grandchildren, or grandparents.

    Note: this condition also nabbed cases where the relationship codes were not correctly identified on one pair member's roster. This occurred rarely, but when it did, the minimum count was 1 and the maximum count was 2.

14. The count of the number of spouse-spouse pairs might not agree because the pair were siblings, but one sibling did not consider a step-parent or parent's partner as a "parent." The maximum count was selected if:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified,
    - The pair members were siblings,
    - The pair member associated with the maximum count had two parents,
    - The pair member associated with the minimum count had one parent, and
    - The roster associated with the pair member with the maximum count had more immediate family members (children, parents, siblings, spouses, partners, grandchildren, or grandparents) than the roster associated with the other pair member.

15. The count of the number of spouse-spouse pairs might not agree because the household otherwise changed after screening, which was not accounted for by previous conditions. In general, the count with a household composition closest to the screener was selected. The age composition was defined by looking at age

classes. The count for a given pair member was selected if the following properties held:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
- The number of roster members between the ages of 26 and 44 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 30 and 49 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 35 and 54 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
- The number of roster members between the ages of 40 and 59 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.

16. In some cases, neither pair member's household composition matched that of the screener. In that case, the household roster closest to that of the screener was selected. The maximum was selected if the following conditions were satisfied:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The number of screener roster members aged 12 or over exceeded the corresponding count from the questionnaire rosters of both pair members, which also differed from each other.

17. If the counts did not match, on the rare occasion one pair member in a spouse-spouse pair identified two grandparents of a different gender. Since there is no code for grandparents-in-law, they could not be identified, so the maximum count was selected. The following conditions were required:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The pair was a spouse-spouse pair.

- The pair member with the maximum count had 2 grandparents of a different gender, and the pair member with the minimum count did not have any.

The assumption here, of course, is that the grandparents of a different gender are in fact a spouse-spouse pair. There is no way to check whether a grandfather is the father's father, and the grandmother is the mother's mother, for example.

18. Even though the household composition may match in terms of ages across the screener roster and the two pair members' rosters, the counts may disagree where two spouse-spouse pairs were clearly identified by one pair member but not the other. This may be because one of the in-laws was incorrectly identified on one side, or because a partner was not considered an in-law by a responding pair member, or because a partner did not consider other family members as "in-laws." The following conditions were required for the maximum count to be selected:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified.
    - The number of screener roster members aged 12 or over matched the corresponding count from the questionnaire rosters of both pair members.
    - The pair member with the maximum number of spouse-spouse pairs had a spouse or partner, and also had two parents.
    - There were no bad relationship codes among roster members 15 or over on either pair member's roster.

19. If the counts for each pair member are not equal, but the number of roster members aged 12 or over is the same between the two pair members, and the count for one pair member is the maximum possible in the household, then that number is selected as the final count. This condition is only applied after all other conditions, including conditions where the final count is ambiguous, have already been applied.

S.4 Spouse-spouse counts (with children)

The household counts for spouse-spouse counts with children obviously depended upon the counts obtained for spouse-spouse counts with or without children. The logic for the spouse-spouse counts with children follows:

1. For a sizable proportion of cases, clearly no couples with children could be in the household, either because the spouse-spouse count was zero, or because the household size was 2 or less. In these cases, the final spouse-spouse-with-children count was set to zero.

2. An additional small number of cases could also be readily determined by looking at the spouse-spouse count. If one pair member had a spouse-spouse with children count that equaled or exceeded the final spouse-spouse count, but the other pair member

had a spouse-spouse with children count which was smaller than the final spouse-spouse count, then the final spouse-spouse with children count was set to the pair member's count that was consistent with the final spouse-spouse count.

For the remainder of general conditions, it had been established that at least one couple resided in the household:

3. For cases that were not already determined by looking at the previous two conditions, the counts for the two pair members (if there were two pair members) were equal in the vast majority of cases. The final count could be set to each pair member's count under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - Two family units did not live in the household that were identified earlier as two family units.
    - Both pair members had valid rosters.
    - Either:
        - The counts were nonzero, and equal to the final spouse-spouse count, or
        - There were no bad relationship codes for roster members under 18, and one of the following held for at least one pair member:
            - The pair member's roster had no bad relationship codes for roster members 15 or over, or
            - The pair member was over 18, and had neither children nor siblings under 18 (covers zero counts, since no bad codes under 18), or
            - The pair member was under 18, did not have parents, but there was one bad relationship code among roster members over 18 in that pair member's roster (covers zero counts, since only one bad relationship code could potentially be a single parent, but not a pair of parents making a couple).

4. The pair members might both have zero counts, but the above conditions did not apply. The final count could still be zero if the age counts for both pair members and the screener indicated nobody lived in the household who was under 18, and there were no bad roster ages.

5. The counts for both pair members might still agree with nonzero counts, even though none of the previous conditions applied. The final count could still be set to one of the pair member's counts if the pair relationship was imputed to be a spouse-spouse pair with children, and there was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had

grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.

6. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:

   - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
   - Either:
     - The count for the pair member with the valid roster was nonzero, and equal to the final spouse-spouse count, or
     - There were no bad relationship codes for roster members under 18, and one of the following held for the pair member with the valid roster:
       - The pair member's roster had no bad relationship codes for roster members 15 or over, or
       - The pair member was over 18, and had neither children nor siblings under 18 (covers zero counts, since no bad codes under 18), or
       - The pair member was under 18, did not have parents, but there was one bad relationship code among roster members over 18 in that pair member's roster (covers zero counts, since only one bad relationship code could potentially be a single parent, but not a pair of parents making a couple).

7. The pair member with the valid roster might have a zero count, but the above conditions did not apply. The final count could still be zero if the age counts for both the pair member with the valid roster and the screener indicated nobody lived in the household who was under 18, and there were no bad roster ages.

8. If the spouse-spouse-with-children counts disagreed in the same manner as the spouse-spouse counts disagreed, then the choice is obvious: use the count that corresponded to the correct spouse-spouse count. Details follow:

   - If the spouse-spouse-with-children counts were equal to the spouse-spouse counts for both pair members, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to the final spouse-spouse count.
   - If the spouse-spouse counts exceeded the spouse-spouse-with-children counts by one for each pair member, even though they were unequal to each other, then the final spouse-spouse-with-children was set to one less than the final spouse-spouse count.

9. If two different family units had already been identified in the household, then two different parent-sets (one often a single parent) were being referenced. The final count was set to the sum of the two counts (where one of the counts was often zero).

10. Based on earlier conditions, we have already excluded households without couples. We have also excluded households with a possibility of two or more couples. If the pair relationship was parent-child, and at least one count was nonzero, then the identified couple must correspond to the parent-child relationship. The maximum of the counts was selected under the following conditions:

    - The sum of counts from the two pair members was 1.
    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - Either:
        - The relationship was parent-child where the child was between 12 and 17, or
        - The relationship was parent-child where the child was between 18 and 20, and the child had siblings under 18.

11. Two couples have been identified in the household, where the household is multigenerational (one member of the younger couple is in a parent-child relationship with the older couple). If a sibling to the pair member in the younger couple was selected, or if a member of the younger couple was selected who "married into" the family, then he or she was not be able to identify the nephews, nieces, brothers-in-law, or sisters-in-law that could point to an appropriate accounting of all the couples with children, because of the relationship codes that were available. The maximum of the two counts was selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - There were two couples in the household, as identified by the final spouse-spouse count.
    - The difference between the pair members' counts was 1.
    - Either:
        - The pair member with the smaller count had a spouse or partner, and the pair member with the larger count had parents in the household, or

  - The pair member with the smaller count had parents-in-law or children-in-law in the household.

12. If a couple is a marriage/partnership that occurred after an earlier marriage, the partner may not consider the partner's children as his or her children, but the child, who was also selected, considered the spouse/partner as a parent. Even though the pair relationship is not parent-child, these cases are still counted as spouse-spouse with children since they are the children of one spouse/partner. The maximum count is selected under the following conditions:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - One count was zero and the other count was one.
    - The pair member with the zero count had a spouse or partner.
    - The pair member with the nonzero count had parents.

13. The counts may have been unequal because children under 18 left, entered, or otherwise materialized or disappeared in the household after screening and between the time of the interviews. In general, the count was selected that corresponded to the pair member with a household composition closest to the screener household composition. If one pair member did not have children in the household, and the other pair member did, the following conditions were required for the count corresponding to the pair member with a household composition closest to the screener:

    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
    - One pair member had a nonzero count of children under 18, and the other pair member had a zero count of children under 18.
    - Either:
      - The screener composition indicated children under 18 were in the household, whereupon the nonzero count was selected, or
      - The screener composition indicated no children under 18 were in the household, whereupon the zero count was selected.

14. The counts may have been unequal with a zero count and a count of one because a pair member with a zero count was not part of the immediate family unit. The nonzero count was used under the following conditions:

- There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.
- The pair relationship was not a parent-child, sibling-sibling, spouse-spouse, or grandparent-grandchild relationship.
- Both pair members had relationship codes that were not parent, child, sibling, spouse, partner, grandparent, or grandchild codes, among roster members who were 12 or over.

The following additional requirement was included, which overly restricted the cases that could be included within this general condition:

- The pair member with a nonzero count was under 21, and had 2 parents.

15. The counts may have been unequal because of bad relationship codes among roster members under 18. The following rules were used to determine if the count associated with the pair member who did not have bad relationship codes:

    - The number of roster members under 18 was the same between both pair members.
    - The side with the smaller count had one bad relationship code for roster members under 18.
    - There was no potential for 2 or more couples in the household that were not already obviously identified, whereby one of the pair members had grandchildren and there were respondents 12 or over who were not children, grandchildren, siblings, children, parents, spouses, or partners, but no children-in-law were identified, and the number of household members was at least 4.

16. If, after considering all of the general conditions given above, the count was left to imputation, it was still possible that the lower and upper bounds were equal. In this instance, the final count was set to one of the bounds.