2005 NATIONAL SURVEY ON DRUG USE AND HEALTH

QUESTIONNAIRE DWELLING UNIT-LEVEL AND PERSON PAIR-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2005 Methodological Resource Book

RTI Project No. 0209009.175.002
Contract No. 283-2004-00022
Deliverable No. 39

Authors:

Project Director: Thomas G. Virag

Matthew Westlake
Jeremy Aldworth
Kortnee Barnett-Walker
Elizabeth Copello
Patrick Chen
Harper Gordek
Jeff Laufenberg

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2007

2005 NATIONAL SURVEY ON DRUG USE AND HEALTH

QUESTIONNAIRE DWELLING UNIT-LEVEL AND PERSON PAIR-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2005 Methodological Resource Book

RTI Project No. 0209009.175.002
Contract No. 283-2004-00022
Deliverable No. 39

Authors:

Matthew Westlake
Jeremy Aldworth
Kortnee Barnett-Walker
Elizabeth Copello
Patrick Chen
Harper Gordek
Jeff Laufenberg

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2007

Preface

This report documents the method of weight calibration used for producing the final set of questionnaire dwelling unit (QDU) and pair weights for the National Survey of Drug Use and Health (NSDUH) data from 2005. The weighting team faced several challenges in this task and was able to address them by resorting to innovative modifications of certain basic statistical ideas. These are listed below.

- Under Brewer's method, high weights may occur due to small pair selection probabilities. In any calibration exercise, some treatment of extreme value (ev) in weights is needed, but there is a danger of introducing too much bias by over-treatment. In the generalized exponential model (GEM), which is described in detail in Chen, et al. (2007), extreme value control is built in, but one needs to define suitable ev domains so that not too many evs are defined. If too many design variables are used to define ev domains, then each domain will be very sparse and will not be of much use in defining thresholds for ev. As in past surveys, a hierarchy of domains was defined using pair age (each pair member being in one of the three categories: 12 to 25, 26 to 49, and 50+) and number of persons aged 12 to 25 in the household, State, and clusters of States (see Section 5.2 for details).

- Control of extreme values in weights helps reduce instability of estimates to some extent, but there is a need for methods that do not introduce much bias. Following the famous suggestion of Hajek (1971) in his comments on Basu's fabled example of circus elephants, we performed ratio adjustment (a form of poststratification) to estimated totals obtained from the household data on the number of persons belonging to the pair domain of interest. This was implemented in a multivariate manner to get one set of final weights.

- In the absence of a suitable source of poststratification controls for the person pair-level weights and the household-level weights, the inherent two-phase nature of the survey design was capitalized upon to estimate these controls from the first phase of the large screener sample. The first-phase sample weight was poststratified to person-level U.S. Bureau of the Census counts to get more efficient estimated counts for pair and household data.

- The problem of multiplicities complicated the issue of providing one set of final weights. When dealing with person-level parameters involving drug-related behaviors among members of the same household, it is possible for an individual to manifest himself or herself in the pair sample through different pairs. To avoid overcounting, the pair weights have to be divided by multiplicity factors, which tend to be domain specific. For this reason, multiplicity factors for a key set of pair analysis domains also are produced along with a set of final calibrated pair weights.

- Missing items in the respondent questionnaire led to imputation for deriving pair relationships, multiplicity factors, and household counts for Hajek adjustments. The general method of predictive mean neighborhood (PMN) was used for this purpose with suitable modifications.

This task required enduring efforts by a dedicated team consisting of Patrick Chen, Matt Westlake, Jeff Laufenberg, and Harper Gordek for weight calibration and Jeremy Aldworth, Elizabeth Copello, and Kortnee Barnett-Walker for imputation. The calibration task described in this document has been in place, with minor modifications, since the 1999 version of NSDUH, which was then called the National Household Survey on Drug Abuse (NHSDA).[1] Results from this calibration applied to an earlier survey year were presented at the 2001 Joint Statistical Meetings. The procedures described in the proceedings papers from these presentations can serve as useful supplemental reference material (Chromy and Singh [2001] on estimation in the presence of multiplicities and extreme weights; Singh, Grau, and Folsom [2001] on the use of PMN; and Penne, Chen, and Singh [2001] on GEM calibration of pair weights). The experience of using GEM with person weights is described in an earlier proceedings paper (Chen, Penne, & Singh, 2000). This work was completed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Office of Applied Studies (OAS), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. 283-2004-00022. The authors would like to take this opportunity to thank a number of individuals for useful discussions and suggestions: Doug Wright, Joe Gfroerer, and Art Hughes of SAMHSA and Jim Chromy and Ralph Folsom of RTI.

NSDUH Weighting Team
Ralph Folsom, Senior Advisor
Research Triangle Park, NC

[1] The National Household Survey on Drug Abuse (NHSDA) was renamed the National Survey on Drug Use and Health (NSDUH) in the 2002 survey year.

Table of Contents

Chapter **Page**

Table of Contents (continued)

Chapter **Page**

Table of Contents (continued)

List of Tables

List of Tables (continued)

List of Tables (continued)

List of Tables (continued)

Table **Page**

List of Exhibits

List of Exhibits (continued)

Exhibit **Page**

List of Terms and Abbreviations

DU	Dwelling unit.
ev	Extreme value. See Sections 5.1 and 5.2 for more detail.
GEM	Generalized exponential model. See Chapter 3 for more detail.
Half step	This refers to halving the increment in the Newton-Raphson iterative process for fitting GEM.
Household-level Person count	The number of pairs associated with a given domain in a given household. These counts are used as control totals in the poststratification step. See Chapter 6 for details on how these counts are created, and Chapter 4 for details on their use in poststratification.
IQR	Interquartile range.
Multiplicity factor	The number of pairs associated with a given respondent in a given domain. See Chapter 6 for more detail.
nr	Nonresponse.
Outwinsor	The proportion of weights trimmed after extreme value adjustment via winsorization.
Pair domain	A pair relationship where the target population is defined by one of the pair members, conditional on the attributes of the other pair member.
Pair relationship	The relationship between selected pair members.
Parent-child	A pair relationship where either both pair members identify the other as part of a parent-child relationship, or both pair members otherwise are determined to form a parent-child pair (either through other evidence or through imputation).
PMN	Predictive mean neighborhood.
ps	Poststratification.
QDU	Questionnaire dwelling unit: a household where at least one member responded to the questionnaire.
res.pr.nr	Respondent pair nonresponse adjustment step. See Section 7.3.3 for more detail.
res.qdu.nr	Respondent questionnaire dwelling unit nonresponse adjustment step. See Section 7.2.3 for more detail.
res.pr.ev	Respondent pair extreme value adjustment step. See Section 7.3.5 for more detail.
res.qdu.ev	Respondent questionnaire dwelling unit extreme value adjustment step. See Section 7.2.5 for more detail.
res.pr.ps	Respondent pair poststratification adjustment step. See Section 7.3.4 for more detail.

List of Terms and Abbreviations (continued)

res.qdu.ps	Respondent questionnaire dwelling unit poststratification adjustment step. See Section 7.2.4 for more detail.
SDU	Screener dwelling unit: a household where screener information is available.
SE	Standard error.
sel.pr.ps	Selected person pair poststratification adjustment step. See Section 7.3.2 for more detail.
sel.qdu.ps	Selected questionnaire dwelling unit poststratification adjustment step. See Section 7.2.2 for more detail.
Sibling-sibling	A pair relationship where the pair members are siblings (either reported to be so, or otherwise determined to be so).
Spouse-spouse	A pair relationship where the pair members are either married or living together as though married (either reported to be so, or otherwise determined to be so).
SS	State sampling.
UWE	Unequal weighting effect. It refers to the contribution in the design effect due to unequal selection probability and is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights and n is the sample size.
Winsorization	A method of extreme value adjustment that replaces extreme values with the critical values used for defining low and high extreme values.

1. Introduction

Traditionally, most household surveys have been designed either to measure characteristics of the entire household or to focus on a randomly selected respondent from among those determined to be eligible for the survey. Selecting more than one person from the same household is considered ill advised since persons from the same household tend to repeat the same general information characteristic of the entire household. Selecting only one person per household avoids the clustering effect on the variance. The "one person per household" sampling approach, however, precludes the opportunity to gather information about the relationships among household members. In the National Survey on Drug Use and Health (NSDUH),[2] we allow for a richer analytic capability of a survey designed to ensure a positive pairwise probability of selection among all eligible household members in each sample household. Achieving positive probabilities for all pairs within sampled households permits unbiased estimation of the within-dwelling-unit component of variance. Besides providing efficient data collection, this sampling method also facilitates the study of the relationships of social behaviors among members of the same household. This report documents the methodology and development of calibrated weights for the second objective, the study of behavioral relationships among persons residing in the same household. The report also describes the development of questionnaire dwelling unit (QDU) weights, which are of independent interest for studying household-level characteristics and also are needed for producing household count estimates of the number of persons belonging to pair relationship domains for use as poststratification controls for pair weights.

NSDUH allows for estimating characteristics at the person level, pair level, and household or QDU level. This report describes the weight calibration methods used for the pair- and QDU-level respondents. As described in the person-level report, NSDUH is an annual survey of about 70,000 persons selected from the civilian noninstitutional population aged 12 or older from all 50 States and the District of Columbia. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. The smaller States were partitioned into 12 State sampling (SS) regions, whereas the eight large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions. Under a stratified design with States serving as the primary strata and SS regions serving as the secondary strata, census tracts, segments, and dwelling units (DUs) for the screener questionnaire were selected using probability proportional to size sampling in two stages. A large number of screener dwelling units (SDUs, about 200,000) were selected to ensure that various age groups (five in all: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50+) of eligible individuals were represented adequately in the final sample. Information collected from SDUs also provided estimates of population controls (as in two-phase sampling) for calibration at levels (such as pair and QDU) for which suitable U.S. Bureau of the Census-based controls were not available. From each selected SDU, zero, one, or two persons were selected using a modification of Brewer's method such that prescribed sampling rates for the five age groups by each State were achieved with high selection rates for youths (12 to 17) and young adults (18 to 25). Table 1.1 shows the eligible number of selected and responding SDUs, QDUs,

[2] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

pairs, and persons for each of the 5 years (2001–2005). The distribution of pair data for different pairs of age groups may vary considerably (see Chapter 2 for details). It is seen that for certain age group domains, the realized sample size may not be sufficient to yield reliable estimates. Also, there may be problems of extreme weights due to small pair selection probabilities under Brewer's method that may cause instability of estimates. These and some other estimation issues related to pair data are discussed below, along with some solutions that were adopted.

Table 1.1 2001 NHSDA and 2002, 2003, 2004, and 2005 NSDUH Sample Sizes

Sample Unit		**2001**	**2002**	**2003**	**2004**	**2005**
SDU	Selected	171,519	150,162	143,485	142,612	146,912
	Completed	157,471	136,349	130,602	130,130	134,055
QDU	Selected	66,697	55,686	56,184	56,251	57,243
	Completed	53,134	48,088	47,753	47,651	47,893
Pair	Selected	23,048	24,895	25,447	25,722	26,562
	Completed	15,795	20,038	20,031	20,109	20,415
Person	Selected	89,745	80,581	81,631	81,973	83,805
	Completed	68,929	68,126	67,784	67,760	68,308

First, we note that for studying drug-related behavioral relationships among members of the same household, pair data is required because the outcome variable generally is defined with respect to the specific other member selected from the household. However, the parameter of interest is generally at the person level and is not at the pair level. For example, in the parent-child pairs, one may be interested in the proportion of children that have used drugs in the past year who have parents that report talking to their child about drugs. Here the target population consists only of children, and not all possible pairs. Note that the pair-level (two persons per QDU) sample forms a subsample of the larger person-level (one or two persons per QDU) sample, with the QDUs themselves selected from the larger sample of SDUs. NSDUH has features of a two-phase design, which turns out to be useful for estimating calibration controls for poststratification of household-level weights and person pair-level weights. No other outside source is available for obtaining these controls. For this purpose, the screener-level household weights are poststratified to person-level census counts to obtain more efficient estimated controls for pair and household data.

In estimation for pair domains, two major problems arise: one is that of multiplicities because, for a given domain defined by the pair relationship, when the parameter of interest is at the person level, several pairs in the household could be associated with the same person. The other problem is that of extreme weights that may arise due to small selection probabilities for certain pair age groups, which may lead to unstable estimates. Each of these issues is discussed in turn.

If several pairs in the household are associated with the same person, it is necessary to use the average measure of behavior relationships for each member, which gives rise to multiplicities. Thus, the pair weights need to be divided by the person-level multiplicity factors for each domain of interest, and, therefore, multiplicity factors need to be produced along with the final set of calibrated weights. Because it is not straightforward to create these multiplicities, analyses would have to be necessarily limited to pair relationships where the multiplicities were produced a priori. It was anticipated that analyses of interest would be limited to 14 pair domains, listed in Table 1.2. Since no multiplicity was necessary for the spouse-spouse/partner-partner pair relationships (by definition, each pair member could have only one partner or one spouse), multiplicity factors were produced for only 12 of these domains. Note that a single pair relationship might have two domains associated with it, since the parameter of interest might be associated with only one member of the pair (the "focus" member), and the multiplicity would differ depending upon which pair member was the focus member.

Table 1.2 Pair Domains

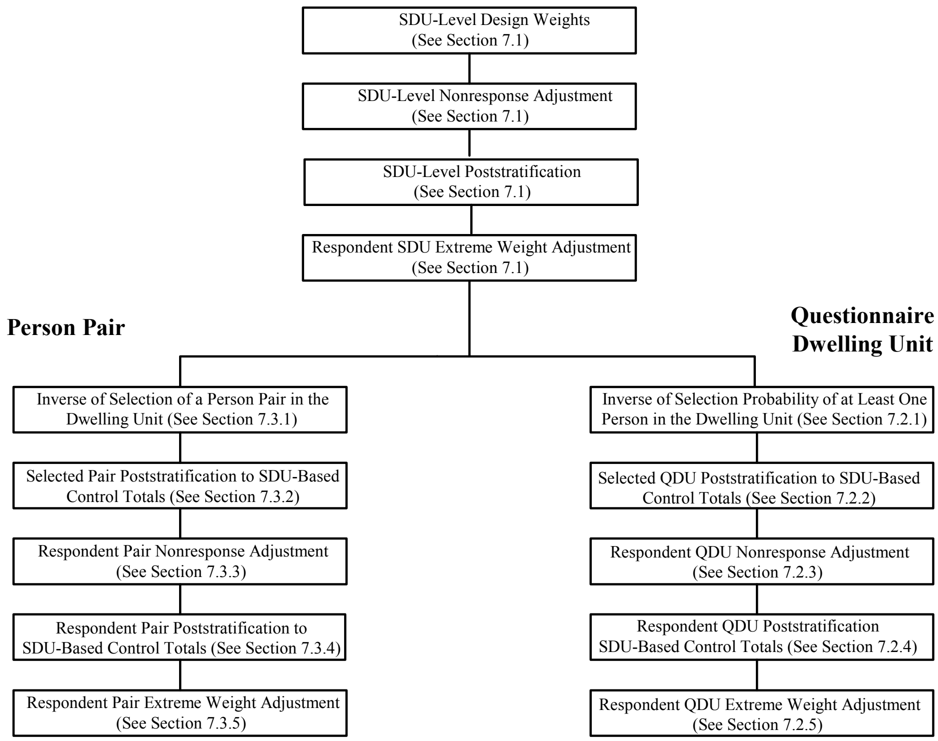
Pair Relationship	Focus
Parent-child, child aged 12-14	Parent
Parent-child, child aged 12-14	Child
Parent-child, child aged 12-17	Parent
Parent-child, child aged 12-17	Child
Parent-child, child aged 12-20	Parent
Parent-child, child aged 12-20	Child
Parent-child, child aged 15-17	Parent
Parent-child, child aged 15-17	Child
Sibling-sibling, older sibling 15-17, younger sibling 12-14	Older sibling
Sibling-sibling, older sibling 15-17, younger sibling 12-14	Younger sibling
Sibling-sibling, older sibling 18-25, younger sibling 12-17	Older sibling
Sibling-sibling, older sibling 18-25, younger sibling 12-17	Younger sibling
Spouse-spouse and partner-partner	No multiplicity necessary
Spouse-spouse and partner-partner, with children aged 0-17	No multiplicity necessary

Some of the multiplicities, including counts of all possible pairs in a household for a given domain, were used for poststratification. Details are provided in Chapter 4.

A resolution to the extreme weight problem is to use a Hajek-type modification (Hajek, 1971). This modification essentially entails calibration (like poststratification) to controls for the number of persons in households belonging to each domain of interest. These controls can be obtained from the larger sample of singles and pairs (i.e., one or two persons selected from DUs). Note, however, that the multiplicity factor, being domain specific, renders the calibration adjustment factor domain specific. This raises the question of finding one set of calibration weights for use with all domains or outcome variables. To get around this problem, we performed a multivariate calibration with respect to a key set of pair domains. This type of poststratification then was followed by a repeat poststratification to further control the extreme weights by imposing separate bound restrictions on the initially identified extreme weights.

The generalized exponential model (GEM) method (Folsom & Singh, 2000) was used for calibration of both QDU- and pair-level design weights through several steps of adjustment as shown in Exhibit 1.1. In GEM, treatment of extreme value (ev) weights is built in via the definition of lower and upper bounds for the extreme weights. For pair data, there was a problem defining suitable domains for defining extreme weights, as explained in the following.

Exhibit 1.1 QDU and Pair Sampling Weight Calibration Steps



In dealing with extreme weights, it is assumed that they arise due to design (due to an imperfect frame, assignment of very small selection probabilities to some units, or a small weight adjustment factor after calibration) so that they do make the sample a representative of population and, hence, do not introduce bias. The only problem is that they may lead to highly unstable estimates similar to the problem of Basu's circus elephants[3] (Hajek, 1971). So, we need

[3] A circus owner had 50 elephants, and wanted to estimate the total weight to help him make arrangements for shipping. To save time, he only wanted to weigh Sambo (an average sized elephant), and use 50 times its weight as an estimate. However, the circus statistician, being highly conscious of the optimality and unbiasedness of the Horvitz-Thompson (HT) estimator, objected about the potential bias of his estimate because of the purposive selection. Instead, he suggested random selection of an elephant with a very high probability of 99/100 for Sambo, and the rest including Jumbo (the biggest in the herd) with probability 1/4900 each. The circus owner was very unhappy with the statistician's response of 100/99 times the Sambo's weight as the estimate if Sambo got selected in this random draw, and was outraged with the response of 4900 times the Jumbo's weight if Jumbo happened to get selected. It was obvious to the owner that this new estimator was extremely poor, although he didn't know anything about its unbiasedness. The story had an unhappy ending with the circus statistician losing his job. To alleviate the instability of the HT-estimator, Hajek suggested to multiply it by 50 divided by inverse of the selection probability, which reduces simply to 50 times the weight of the selected elephant.

to perform some treatment (such as winsorization[4]) within suitably defined extreme weight domains such that these domains contain units possibly from different strata but with similar sample selection probabilities to avoid the occurrence of extreme weights due to a mix of different designs. The domains must be large enough (say, at least size 30) to be able to define extreme values according to the domain-specific weight distribution. Any extreme value treatment to increase precision of estimates would introduce some bias. However, this bias can be reduced considerably if the ev treatment is performed under calibration controls. This is what the built-in ev control in GEM tries to accomplish.

It follows that the definition of extreme weight domains should depend on factors that affect the selection probabilities of units in the sample, such as State- and age-specific sampling rates, segment selection probabilities, pair age-specific selection probabilities, and household composition. If one tries to define extreme weight domains by taking account of all these factors via cross-classification, it will lead to too many domains with insufficient observations. That is why it is difficult to define suitable extreme weight domains for pair data. In the case of person-level weights it was less difficult, since State by age group suitably captured the extreme weight domain requirements. The definition of extreme weight domains used in the 2005 survey was the same as the one used in the 1999–2004 surveys. The domains were defined as the cross-classification of State, pair age,[5] and number of persons aged 12 to 25 in a household. In particular, the pair age was defined by the age groups of each pair member according to the age categories of 12 to 25, 26 to 49, and 50 or older (resulting in six pair age categories), and the number of persons aged 12 to 25 were categorically defined as zero, one, and two or more. For more details, see Chapter 5.

---

[4] A method of extreme value adjustment that replaces extreme values with the critical values used for defining low and high extreme values.

[5] Not to be confused with the modeling term, which has a finer level breakdown.

2. Questionnaire Dwelling Unit and Pair Selection Probabilities

Similar to the 1999–2001 National Household Surveys on Drug Abuse (NHSDAs) and the 2002, 2003, and 2004 National Surveys on Drug Use and Health (NSDUHs),[6] the 2005 NSDUH had a two-phase design and used a computer-assisted interviewing (CAI) method. There were four stages of selection: census tracts, segments within census tracts, dwelling units (DUs) within segments, and persons within dwelling units. Any two survey eligible persons had some nonzero chance of being selected and, when both were selected, they formed a within household pair. This design feature is of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of an individual (in a family) relates to the drug use propensity of other members residing in the same dwelling unit (Morton, Chromy, Hunter, & Martin, 2006).

For the 1999–2001 surveys, the method used for selecting pairs was as follows. For a given DU, if the sum of the age-specific selection probabilities was larger than 2, then the individual person-selection probabilities were ratio adjusted downward to make the sum equal to 2. If the sum was less than 2, the difference between 2 and the sum of the probabilities was evenly distributed over 3 dummy persons so that the sum of the person probabilities was made to equal 2. Brewer's method then was applied to select a person pair using the pair selection formula (2.1). If the selected pair consisted of two real persons, then both persons were selected. If the selected pair consisted of one real person and one dummy person, then the real person was selected. If the selected pair consisted of two dummy persons, no one was selected from that DU.

The 2005 survey pair-sampling algorithm was the same as that used in 2002, 2003, and 2004, which was modified from the one used for the previous surveys (1999–2001) to increase the number of pairs selected in the sample, as explained below. For dwelling units with the sum of person-level selection probabilities (denoted by S) less than 2, the earlier algorithm was modified (see Case II below) to increase the chance for selecting a pair. However, for dwelling units with $S \geq 2$, there was no need to change the algorithm denoted below by Case I. A summary of sample size by pair age and other domains is provided in Table 2.1.

2.1 Pair Selection Probability

Case I: DUs with $S \geq 2$

For a given DU, if the sum of the age-specific person selection probabilities (S) was larger than 2, then the selection probability was ratio adjusted by a multiplicative adjustment factor so that all probabilities were scaled down to sum to exactly 2. Now, Brewer's method sets the pairwise selection probabilities at

[6] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

$$P_{h(ij)} = \left[\frac{P_{h(i)}P_{h(j)}}{K}\right]\left[\frac{1}{1-P_{h(i)}} + \frac{1}{1-P_{h(j)}}\right] \tag{2.1}$$

by setting K at

$$K = 2 + \sum \frac{P_{h(i)}}{1-P_{h(i)}}, \tag{2.2}$$

where i = i^{th} person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5) and
j = j^{th} person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5),

where age category 0 corresponds to dummy persons, 1 to persons aged 12 to 17, 2 to persons aged 18 to 25, 3 to persons aged 26 to 34, 4 to persons aged 35 to 49, and 5 to persons aged 50 or older.

The sum of the pairwise selection probabilities taken over all unique pairs will be guaranteed to be exactly 1.

$$\sum_{i}\sum_{j>i} P_{h(ij)} = 1 \tag{2.3}$$

It also guarantees that the sum of the pairwise selection probabilities for an individual is equal to the individual's selection probability

$$\sum_{j \neq i} P_{h(ij)} = P_{h(i)} \tag{2.4}$$

for all values of i.

As noted earlier, the above scheme will always select a pair, but it allows all combinations of eligible persons and dummy persons so that zero, one, or two eligible persons are selected.

Case II: DUs with S < 2

If the sum S of person-level selection probabilities was less than 2, the earlier method used in previous surveys consisted of dividing 2 - S equally among the 3 dummy persons added to the household, and then used Brewer's method, as in Case I. However, if the household had two or more persons, we would have preferred more chances for a pair to be selected. To achieve this goal, the individual selection probabilities, $P_{h(i)}$, were scaled upward by the factor F_s such that their sum came close to but did not exceed 2 and such that each person selection probability did not exceed the maximum allowed probability of 0.99. Thus, denoting the revised person selection probabilities by $P'_{h(i)}$, the factor F_s is given by

$$F_s = Min\left\{\frac{T(\lambda)}{S}, \frac{0.99}{\max\left\{P_{h(i)}\right\}}\right\}, \qquad (2.5)$$

where $T(\lambda) = S + \lambda(2\text{-}S)$ and λ is set to 0.5. Note that if λ is chosen as 0, then $F_s = 1$, as in Case I. The individual person probabilities are scaled upward by the factor F_s to sum to 2 or as close to 2 as possible. If, after scale adjustment, the sum S' is exactly 2, then dummy persons are not needed. If S' is less than 2, 3 dummy persons are added as before.

Now, for Brewer's method, we set the pairwise selection probabilities similar to (2.1), as

$$P_{h(ij)} = \left[\frac{P'_{h(i)}P'_{h(j)}}{K}\right]\left[\frac{1}{1-P'_{h(i)}} + \frac{1}{1-P'_{h(j)}}\right] \qquad (2.6)$$

by setting K' at

$$K' = 2 + \sum_i \frac{P'_{h(i)}}{1-P'_{h(i)}}, \qquad (2.7)$$

where $P'_{h(i)}$ and $P'_{h(j)}$ are the selection probabilities adjusted by the scaling factor (F_s),
where $i = i^{th}$ person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5), and
$j = j^{th}$ person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5).

Note that we now have $\sum_{j \neq i} P'_{h(ij)} = P'_{h(i)}$. To maintain the original person selection probabilities despite the scale adjustment by F_s, we modified the above Brewer's method as follows. First draw a random number R from uniform (0,1). If $R \leq 1/F_s$, then select a pair using Brewer's method based on formula (2.6). However, if $R > 1/F_s$, then no persons are selected from the household. In this way, the probability for selecting a pair (i,j) in household h becomes $P^*_{h(ij)} = P'_{h(ij)}/F_s$, which, in turn, gives the original person selection probabilities, $P_{h(i)}$.

2.2 Questionnaire Dwelling Unit Selection Probability

A dwelling unit was considered a selected questionnaire dwelling unit (QDU) if it had completed the screening interview and had at least one person selected for the questionnaire interview. QDUs with at least one respondent were considered respondent QDUs.

The QDU selection probability was defined as

$$P_h^* = \left(1 - P_{h(00)}^*\right), \tag{2.8}$$

where $P^*_{h(00)}$ is the probability of not selecting any person. For the DUs with the sum of age-specific selection probabilities larger than or equal to 2 (Case I), $P^*_{h(00)}$ is 0. It follows from Section 2.1, under Case II, $P^*_{h(00)}$ can be calculated as

$$P_{h(00)}^* = \left(1 - \frac{1}{F_s}\right) + \frac{3}{F_s}\left[\frac{P'_{h(0)}P'_{h(0)}}{K'}\right]\left[\frac{1}{1-P'_{h(0)}} + \frac{1}{1-P'_{h(0)}}\right], \tag{2.9}$$

where $P'_{h(0)}$ is the selection probability of a dummy person when person selection probabilities are adjusted by F_s.

Table 2.1 Distribution of Pair Age Groups for the 2001 NHSDA Sample Sizes and the 2002, 2003, 2004, and 2005 NSDUH Sample Sizes

	2001			2002			2003			2004			2005		
Domain	**Sel.**[1]	**Resp.**[2]	**% Rate**[3]	**Sel.**[1]	**Resp.**[2]	**% Rate**[3]	**Sel.**[1]	**Resp.**[2]	**% Rate**[3]	**Sel.**[1]	**Resp.**[2]	**% Rate**[3]	**Sel.**[1]	**Resp.**[2]	**% Rate**[3]
DUs															
Total DUs Screened	157,469	n/a	n/a	136,349	n/a	n/a	130,605	n/a	n/a	130,130	n/a	n/a	134,055	n/a	n/a
QDUs															
Total QDUs	66,697	53,134	79.66	55,686	48,088	86.36	56,184	47,753	84.99	56,251	47,651	84.71	57,243	47,893	83.67
Persons															
Total Persons	89,745	68,929	76.81	80,581	68,126	84.54	81,631	67,784	83.04	81,973	67,760	82.66	83,805	68,308	81.51
12-17	28,188	23,133	82.07	26,230	23,645	90.14	25,387	22,665	89.28	25,141	22,301	88.70	25,840	22,534	87.21
18-25	30,304	22,658	74.77	27,216	23,066	84.75	27,259	22,738	83.41	27,408	22,829	83.29	27,337	22,511	82.35
26-34	8,825	6,893	78.11	7,672	6,374	83.08	8,060	6,570	81.51	8,052	6,574	81.64	8,573	6,856	79.97
35-49	13,663	10,036	73.45	12,076	9,620	79.66	12,604	9,831	78.00	12,907	9,951	77.10	13,202	10,099	76.50
50+	8,765	6,209	70.84	7,387	5,421	73.39	8,321	5,980	71.87	8,465	6,105	72.12	8,853	6,308	71.25
Pairs															
Total Pairs[4]	23,048	15,795	68.53	24,895	20,038	80.49	25,447	20,031	78.72	25,722	20,109	78.18	26,562	20,415	76.86
12-17, 12-17	4,772	3,724	78.04	4,667	4,192	89.82	4,302	3,860	89.73	4,292	3,810	88.77	4,476	3,913	87.42
12-17, 18-25	3,534	2,475	70.03	3,245	2,742	84.50	3,206	2,672	83.34	3,206	2,678	83.53	3,283	2,643	80.51
12-17, 26-34	836	604	72.25	826	694	84.02	927	758	81.77	845	699	82.72	879	724	82.37
12-17, 35-49	4,054	2,848	70.25	3,795	3,121	82.24	3,892	3,180	81.71	3,901	3,135	80.36	4,187	3,288	78.53
12-17, 50+	524	341	65.08	482	377	78.22	584	446	76.37	603	435	72.14	623	453	72.71
18-25, 18-25	5,921	3,716	62.76	5,520	4,419	80.05	5,522	4,285	77.60	5,642	4,426	78.45	5,508	4,255	77.25
18-25, 26-34	875	602	68.80	975	806	82.67	1,085	848	78.16	1,073	860	80.15	1,206	900	74.63
18-25, 35-49	1,329	792	59.59	1,449	1,042	71.91	1,533	1,079	70.38	1,578	1,072	67.93	1,609	1,119	69.55
18-25, 50+	574	301	52.44	604	418	69.21	695	453	65.18	726	471	64.88	763	509	66.71
26-34, 26-34	177	129	72.88	774	559	72.22	856	623	72.78	849	615	72.44	957	688	71.89
26-34, 35-49	111	72	64.86	450	346	76.89	493	372	75.46	490	356	72.65	563	382	67.85
26-34, 50+	41	31	75.61	196	123	62.76	195	113	57.95	233	135	57.94	220	138	62.73
35-49, 35-49	152	85	55.92	807	543	67.29	892	579	64.91	954	601	63.00	867	565	65.17
35-49, 50+	63	29	46.03	350	210	60.00	392	228	58.16	410	228	55.61	447	254	56.82
50+, 50+	85	46	54.12	755	446	59.07	873	535	61.28	920	588	63.91	974	584	59.96

DU = dwelling unit; QDU = questionnaire dwelling unit.

[1] Selected pairs are based on the screener age.

[2] Respondent pairs are based on questionnaire age.

[3] These rates are unweighted and based only on the total selected and total responding counts of pairs.

[4] Total pairs excludes dummy, person pairs.

3. Brief Description of the Generalized Exponential Model

In survey practice, design-based weights are typically adjusted in three steps: (1) for extreme values (ev) via winsorization, (2) for nonresponse (nr) adjustment via weighting class, and (3) for poststratification (ps) via raking ratio adjustments. If weights are not treated for extreme values, the resulting estimates, although unbiased, will tend to have low precision. The bias introduced by winsorization is alleviated to some extent through ps. The nr adjustment is a correction for bias introduced in estimates based only on responding units, and ps is an adjustment for coverage (typically undercoverage) bias and variance reduction due to correlation between the study and control (usually demographic) variables.

There are limitations in the existing methods of weight adjustment for ev, nr, and ps. It would be desirable to adjust for bias introduced in the ev step (when extreme weights are treated via winsorization) in that the sample distribution for various demographic characteristics is preserved. For the nr step, there are general raking type methods, such as the scaled constrained exponential model developed by Folsom and Witt (1994), where the lower and upper bounds can be suitably chosen by use of a separate scaling factor. The factor is set as the inverse of the overall response propensity. It would be desirable to have a model for the nr adjustment factor so that the desired lower and upper bounds on the factor are part of the model. Note that the lower bound on the nr adjustment factor should be one, as it is interpreted as the inverse of the probability of response for a particular unit. For the ps step, on the other hand, the general calibration methods of Deville and Särndal (1992), such as the logit method, allow for built-in lower (L) and upper (U) bounds (for ps, typically $L < 1 < U$). However, it would be desirable to have nonuniform bounds (L_k, U_k) depending on the unit k such that the final adjusted weight, w_k, could be controlled within certain limits. An important application of this feature would be weight adjustments in the presence of ev such that the user will have some control on the final adjustment of the initially identified extreme values.

A modification of the earlier method of the scaled constrained exponential model of Folsom and Witt (1994), termed as the method of the generalized exponential model (GEM) and proposed by Folsom and Singh (2000), provides a unified approach to the three weight adjustments for ev, nr, and ps, and it has the desired features mentioned above. The functional form of the GEM adjustment factor is provided in Appendix A. It generalizes the logit model of Deville and Särndal (1992), typically used for ps, such that the bounds (L, U) may depend on k. Thus, it provides a built-in control on ev during both ps and nr adjustments. In addition, the bounds are internal to the model and can be set to chosen values (e.g., $L_k = 1$ in the nr step). If there is a low frequency of ev in the final ps, then a separate ev step may not be necessary.

In fitting GEM to a particular problem, the choice of a large number of predictor variables along with tight bounds will have an impact on the resulting unequal weighting effect (UWE) and the proportion of extreme values. In practice, this leads to somewhat subjective considerations of trade-off between the target set of bounds for a given set of factor effects and the target UWE and the target proportion of extreme values. It also may be beneficial to look at the proportion of "outwinsors" (a term coined to signify the extent of residual weights after winsorization), which is probably more realistic in determining the robustness of estimates in the presence of extreme values.

A large increase in the number of predictor variables in GEM typically would result in a higher UWE, thus indicating a possible loss in precision. This was checked by comparing SUDAAN-based standard errors of a key set of estimates computed from two sets of calibration models, one baseline using only the main effects and the other using the final model. The results are presented in Chapter 8.

To implement GEM, several steps need to be followed: (1) define and create all the covariates; (2) define the extreme weights; (3) fit the GEM model. The details of practical aspects of GEM implementation can be found in Chapters 4 and 5 of this report and Chapter 4 of Chen et al. (2007).

4. Predictor Variables for the Questionnaire Dwelling Unit and Pair Weight Calibration via the Generalized Exponential Model

We note that unlike the person-level weight calibration, the control totals for the questionnaire dwelling unit (QDU)-level and person pair-level poststratification are not available from the U.S. Bureau of the Census. A way around this potential problem is to take advantage of the two-phase nature of the design, in which the screener data provides a large sample containing demographic information that can be used to derive control totals for the QDU-level and person pair-level sampling weight calibrations, as well as for the selected person poststratification adjustment. The stability of control totals from the screener dwelling unit (SDU)-level data can be improved by poststratification of the SDU sample using person-level counts from the census. This was indeed done and is documented in the person-level weight calibration report (Chen et al., 2007).

4.1 QDU Weight Calibration

After the nonresponse and poststratification adjustments at the SDU level, which are common to the person-level weight calibration, the QDU sample weights were adjusted in three steps: poststratification of selected QDUs, nonresponse adjustment of respondent QDUs, and poststratification of respondent QDUs. The set of initially proposed predictor variables for these adjustments using generalized exponential model (GEM) were set to be common and to correspond to those used for the SDU nonresponse and poststratification adjustments. The variables are of two types: Those used for SDU nonresponse adjustment are 0/1 indicators, while those used for SDU poststratification adjustment are counting variables. The variables of the first type (0/1 indicators) are population density, group quarters, race/ethnicity of householder, percentage of persons in segment who are black or African American, percentage of persons in segment who are Hispanic or Latino, percentage of owner-occupied dwelling units (DUs) in segment, segment-combined median rent and housing value, and household type. Variables of the second type (counting variables) represent the number of eligible persons within each DU who fall into the various demographic categories of race, age group, Hispanicity, and gender. Note that the State and quarter variables are represented as both binary and counting variables. Thus, not only are DU counts within a specific State or quarter in the QDU sample controlled to the corresponding totals obtained from the SDU sample, but also counts of persons living in the DUs in the QDU sample are controlled to totals from the SDU sample. These person-level totals match the census estimates because of the SDU-level poststratification to census counts. It may be noted that in the poststratification of selected QDUs and the nonresponse adjustment of the respondent QDUs steps, demographic information from screener data was used in defining covariates, whereas in the poststratification of the selected QDUs step, questionnaire demographic information was used.

Exhibit 4.1 lists all predictor variables proposed for QDU-level calibration and identifies them as counting, binary, or both. Various main effects and higher level factor effects based on the predictor variables were included in the GEM modeling. As stated previously, all adjustment steps at the QDU level used a common set of proposed predictor variables.

4.2 Pair Weight Calibration

Like QDU, the initial set of weight components in pair weight calibration are the same as the set obtained from the SDU-level weight calibration. The SDU-calibrated weight is multiplied by the pair-level design weight, which in turn was adjusted in four steps: poststratification of selected pairs, nonresponse adjustment of respondent pairs, poststratification of respondent pairs, and the extreme weight adjustment of respondent pairs. All the adjustment steps for pair weights utilized the same set of initially proposed predictor variables, which included a subset of those used for the person-level nonresponse adjustment. This included segment characteristic variables, such as population density, percentage of persons in segment who are black or African American, percentage of persons in segment who are Hispanic or Latino, percentage of owner-occupied DUs in segment, and segment-combined median rent and housing value. Also included were pair-specific covariates, such as the demographic characteristics of pair age, pair race/ethnicity, and pair gender, as well as dwelling unit characteristics, such as race/ethnicity of householder, household type, household size, and group quarters indicators. State and quarter indicators were included as well. However, for two-factor effects, instead of individual State, State/region was used due to insufficient sample size. This resulted in a 12-level variable where the eight large sample States were kept separate, and the remainder of States were grouped according to the four census regions. All variables were defined as 0/1 indicators. These proposed predictor variables and their levels are shown in Exhibit 4.2.

In the poststratification of selected pairs and the nonresponse adjustment of respondent pairs, screener data were used in the definition of the pair-specific variables such as pair age, pair race/ethnicity, and pair gender, whereas in the poststratification and extreme weight adjustment of respondent pairs, these variables were obtained from the questionnaire. For the latter case, in addition to the variables described above, indicator covariates corresponding to selected pair domains were included to perform Hajek-type ratio adjustments via weight calibration, as mentioned in Chapter 1. The selected pair domains were limited to 10 of the 14 pair domains listed in Chapter 1. (Parent-child pairs where the child was in the 15- to 17-year-old age range and sibling-sibling-younger sibling focus pairs were not included in the poststratification.) The inclusion of these pair domain covariates led to the use of two sets of control totals in the modeling. Details of the construction of these control totals can be found in Appendix B.

Exhibit 4.1 Definitions of Levels for QDU-Level Calibration Modeling Variables

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

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

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

Hispanicity[b]
 1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]

Household Size[b]
 Continuous variable count of individuals rostered with DU

Household Type (Ages of Persons Rostered within DU)[a]
 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+[1]

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

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

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

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

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

Race (3 Levels)[b]
 1: white,[1] 2: black or African American, 3: other

Race (5 Levels)[b]
 1: white,[1] 2: black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: multirace

Race/Ethnicity of Householder[a]
 1: Hispanic or Latino white,[1] 2: Hispanic or Latino black or African American, 3: Hispanic or Latino others, 4: Non-Hispanic or Latino white, 5: Non-Hispanic or Latino black or African American, 6: Non-Hispanic or Latino others

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

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

States[a,b,3]

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[a,3]
 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]

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.
[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 percentage owner-occupied.
[3] The State or district assigned to a particular model is based on census regions.
[a] Binary variable.
[b] Counting variable.

Exhibit 4.2 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
1: DU with 2 persons,[1] 2: DU with 3 persons, 3: DU with $\geq$ 4 persons

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

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

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

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

Pair Race/Ethnicity (10 Levels)
1: white and white,[1] 2: white and black or African American, 3: white and Hispanic or Latino, 4: white and other, 5: black or African American and black or African American, 6: black or African American and Hispanic or Latino, 7: black or African American and other, 8: Hispanic or Latino and Hispanic or Latino, 9: Hispanic or Latino and other, 10: other and other

Pair Race/Ethnicity (5 Levels)
1: Mixed race pair, 2: Hispanic or Latino pair, 3: black or African-American pair, 4: white pair,[1] 5: other pair

Pair Race/Ethnicity (4 Levels)
1: Mixed race/ethnicity pair or other and other, 2: Hispanic or Latino pair, 3: black or African-American pair, 4: white pair[1]

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

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

Percentage of Segments That Are Hispanic or Latino (% 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/Ethnicity of Householder
1: Hispanic or Latino white,[1] 2: Hispanic or Latino black or African American, 3: Hispanic or Latino others, 4: Non-Hispanic or Latino white, 5: Non-Hispanic or Latino black or African American, 6: Non-Hispanic or Latino others

Exhibit 4.2 Definitions of Levels for Pair-Level Calibration Modeling Variables (continued)

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;[1] 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming; 3: Michigan; 4: Illinois; 5: Ohio; 6: California

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/partner-partner
10: Spouse-spouse/partner-partner with children (younger than 18)

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.
[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 percentage owner-occupied.
[3] The States or district assigned to a particular model is based on combined census regions.
* The pair member focused on.

5. Definition of Extreme Weights

An important feature of the generalized exponential model (GEM) is the built-in provision of extreme value (ev) treatment. For this purpose, sampling weights are classified as extreme (high or low) if they fall outside the interval, median ± 3*interquartile range (IQR), for some prespecified domains defined usually by design variables corresponding to deep stratification.[7] The critical values for low and high extreme values will be denoted by $b_{k(l)}$ and $b_{k(u)}$, respectively. Within GEM modeling, these critical values were defined as median ± 2.5*IQR, which were conservative when compared with the commonly used standard of median ± 3*IQR. This is because in order to better prevent the adjusted weights from crossing the standard boundary, weights near but below it (i.e., those that have the most potential to become extreme) were treated as extreme by GEM as well.

For implementing extreme value control via GEM, the variable m_k was defined as the minimum of $b_{k(u)}/w_k$ and one for high extreme weights, and the maximum of $b_{k(l)}/w_k$ and one for low extreme weights, where w_k represents the sampling weight before adjustment, and $b_{k(u)}$ and $b_{k(l)}$ denote the critical values for the extreme weights. (Note that under this definition, for high extreme weights, the more extreme the weight is, the smaller m_k will be, and, conversely, for low extreme weights, the more extreme the weight is, the bigger m_k will be.) Nonextreme weights had a value of one for m_k. The upper and lower bounds for the adjustment factors were defined, respectively, as the product of m_k and the upper and lower boundary parameters of GEM. GEM allows inputs of up to three different upper and lower boundary parameters (L_1 and U_1, L_2 and U_2, L_3 and U_3) for high, non-, and low extreme weights. By applying a small upper boundary parameter for high extreme weights and a large lower boundary parameter for low extreme weights, the extreme weights can be controlled in the modeling process.

5.1 Questionnaire Dwelling Unit Extreme Weight Definition

For the questionnaire dwelling unit (QDU)-level weight adjustment, extreme weights were defined using a nested hierarchy of six domains:

1. State;
2. State sampling region;
3. State by household type;

Levels of household type indicate whether the household has members who are youths, young adults, or adults, where youth signifies 12 to 17, young adult 18 to 25, and adult 26 or older.

a. Youth, Young Adult, Adult;
b. Youth, Young Adult;
c. Youth, Adult;
d. Young Adult, Adult;

[7] Deep stratification refers to the stratification that was used in the sample design. In the case of the 2005 survey, deep stratification refers to the cross-classification of State sampling region by age group.

   e. Youth Only;
   f. Young Adult Only; and
   g. Adult Only.

4. State sampling region by household type;
5. State by household type by household size (1, 2, 3, 4+); and
6. State sampling region by household type by household size.

The hierarchy is used to satisfy the minimum of 30 observations for defining the boundaries for extreme values. If this sample size requirement is not met at the lower level, then the next level up in the hierarchy is used.

5.2 Person Pair Extreme Weight Definition

The pair selection probability is a function of the selection probability of each person in the pair given by formula (2.1) or (2.6), depending on the sum of the person selection probabilities within the household as discussed in Section 2.1. This probability can be very small if the selection probabilities of individual members are small. For example, consider a selected dwelling unit (DU) (ID = RI59070135) from the 2005 survey. This DU gave rise to a selected pair of respondents, both aged 50 or older. The selection probability for a respondent aged 50 or older was 0.08445. Using the formula (2.6) in Chapter 2, the pair selection probability was computed to be 0.000210. Therefore, the inverse of the probability, the pair-level design weight, was 4,750.92. Thus, a small pair selection probability can create a high initial weight, which is the product of the screener dwelling unit (SDU) weight and the person pair design-based weight.

As mentioned in the introduction, it turns out to be difficult to select suitable domains for defining extreme weights for pair-level data. However, as was done for the 1999–2004 surveys, the extreme weight definition was based on the following hierarchy of domains:

1. Pair age group[8] (with three age categories, 12 to 25, 26 to 49, and 50+) by number (0, 1, 2+) of persons aged 12 to 25 in the household;
2. State cluster (with five levels [explained below]) by pair age group by number (0, 1, 2+) of persons aged 12 to 25 in the household;
3. State cluster (with three levels [explained below]) by pair age group by number (0, 1, 2+) of persons aged 12 to 25 in the household; and
4. State by pair age group by number of persons aged 12 to 25 (0, 1, 2+) in the household.

The hierarchy is used to satisfy the minimum of 30 observations for defining the boundaries for extreme values. If this sample size requirement is not met at the lower level, then the next level up in the hierarchy is used.

We now briefly introduce the considerations behind the above definition for extreme weight domains. The sample design prespecified the person-level selection probability within

[8] Not to be confused with the modeling term, which has a finer level breakdown.

State by five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50+). Age groups 12 to 17 and 18 to 25 have a relatively similar selection probability, and the same is true for age groups 26 to 34 and 35 to 49. The 50+ group, however, has a quite different selection probability from the other groups. Furthermore, since the 12 to 17 and 18 to 25 age groups have large selection probabilities, they have a very high chance of being selected if the household has persons in these age groups. Therefore, the number of persons aged 12 to 25 in the household has a significant impact on the type of pair selected and the pair selection probability. Taking into consideration these design-related features, a suitable domain to define the pair-level extreme weight seems to be given by State by pair age group by number of persons aged 12 to 25 in the household.

The hierarchy of domains mentioned above was used to satisfy the minimum of 30 observations. However, it was found that for many ev domains the minimum sample size requirement was not met. To get around this problem, States were grouped into a small number of clusters, say three or five. The assignment of States to clusters was determined by the clustering algorithm in PROC CLUSTER in SAS, where the clustering variable was defined as the average person-level weight (ANALWT) for each of the five age groups within each State. The choice of the average person-level weight for each group for each State was motivated from the objective of finding a single variable that would reflect the design-based difference in pair selection probabilities across States. Even with clustering of States, the ev domain sample size may be insufficient, so the most general level of the hierarchy, the national level, is required. Even at the national level, we had to collapse some pair age categories in forming domains of reasonable sample size to define extreme weights. More specifically, for the national level, we collapsed all levels of number of persons aged 12 to 25 for the pair age groups of 50+, 50+ and 26 to 49, 50+. In addition, levels 1 and 2+ of number of persons aged 12 to 25 were combined for the pair age group of 26 to 49, 26 to 49.

6. Editing and Imputation of Pair Relationships, Multiplicity Factors, and Household-Level Person Counts for Poststratification

6.1 Introduction

"Pair data" are used to study outcome variables among members of the same household. These outcome variables are measured using the "pair relationship," the relationship between selected pair members. For these analyses, the outcome variables may be at either the person level or the pair level. The most common type of analysis is the person-level analysis, where the inferential population is defined by one of the pair members. This pair member is the "focus" pair member. An example of an outcome at the person level is the proportion of youths who use drugs and whose parents report talking to them about drugs, where the focus is on the youth in a parent-child pair. An example at the pair level is child-parent drug behavior for all possible parent-child pairs (within the youth's age group). Knowledge of the pair relationship and the inferential population gives rise to the "pair domain."

For analyses at the pair level, the pair domain is completely defined by the pair relationship, whereas the pair domain for a person-level analysis depends upon which pair member is the focus. "Multiplicity" is an issue that arises in the analysis of pair data in which the analysis is at the person level for a given pair domain. Several pairs in the household could be associated with the same person. Consider the previous example where we are interested in the proportion of children who use drugs and whose parent reports talking to them about drugs. In this case, if the household has two parents, the selected child has two inclusion possibilities (one with each parent) in the set of all such parent-child pairs. Since children form the target population for this example, it is desirable to assign one observation per child. A reasonable way to achieve this is to take an average of the two responses, which together correspond to the two pairs associated with the child (i.e., one for each parent in this example). In other words, the response for each child-parent pair from two-parent households is divided by the number of parents. This divisor is known as the "multiplicity factor." The multiplicity problem does not arise if there is only one inclusion possibility (e.g., a single-parent household, if the child is the focus) or if the analysis is a pair-level analysis (e.g., child-parent pair drug behavior).

To illustrate how multiplicities appear in the definitions of parameters and estimates, consider estimation of the total number of children who used drugs in the past year, where a parent reported talking to them about drugs. Let $y_{hip}(d)$ be defined as the drug-related behavior outcome for pair p containing the individual i belonging to domain d in household h. Now, for the population of all individuals who belong to the domain d, the total parameter is defined as (Chromy & Singh, 2001)

$$\tau_y(d) = \sum_{h=1}^{H} \sum_{i=1}^{N_h(d)} \sum_{p=1}^{M_{hi}(d)} \frac{y_{hip}(d)}{M_{hi}(d)}$$

(i.e., total of averages over pairs (p) associated with the individual i, over all i in domain d and in the household h). Here $M_{hi}(d)$ denotes the multiplicity (i.e., the number of pairs associated) with the person i in domain d, and $N_h(d)$ can be thought of as the multiplicity count for the household h (i.e., the number of persons in the household that are in domain d). This latter multiplicity count is equivalent to the household-level person count described in the next paragraph. For the sake of simplicity, the weights are not shown in the above estimator.

The predictor variables used for all previous generalized exponential model (GEM)[9] adjustment steps also were used in the respondent pair poststratification (ps) step. In addition to these variables, 10 covariates derived from 10 pair domains were included in the weight adjustment process. This was done to obtain more stable pair-level analysis weights. (A total of 14 pair domains were identified for the purposes of creating multiplicities, as discussed in Section 6.3. Ten of those were used to poststratify the pair-level analysis weights, as discussed in Section 6.4.) Each covariate was defined by the appropriate pair relationship divided by its associated multiplicity. In this ps step, for these 10 pair domains, the nonresponse (nr)-adjusted weights were poststratified to the final questionnaire dwelling unit (QDU) weights. The household-level person counts, which are counts of the number of pairs in the household belonging to a given pair domain, were used to form the control totals in the ps step for these domains. For other domains, the control totals were formed by the screener dwelling unit (SDU) weights from all the possible screener pairs associated with the number of possible pairs in the dwelling unit.

In the process of setting up variables for analyses at the pair level, three types of variables, which are not weights, required editing and imputation. The procedures associated with these three types of variables are referred to as stages. Stage one refers to the creation and imputation of the variables that identify the pair relationships. The multiplicity and household-level person counts that are described in the preceding paragraph were created and imputed in stages two and three, respectively. Missing values in all three stages were imputed using the semiparametric predictive mean neighborhood (PMN) imputation procedure, which uses predicted means from models to find donors in a nearest neighbor hot deck. The hot deck is described in Appendix M, and the PMN procedure is described in detail in Appendix N.

6.2 Stage One: Creation and Imputation of Pair Relationships

6.2.1 Editing the Household Roster of Each Pair Member

Prior to the identification of the relationships between selected pair members, a key step was to edit the questionnaire household rosters for each pair member. This involved identifying situations where the relationship listed in the roster for a particular roster member was not possible given the roster member's age and relationship to the respondent. In many cases, this resulted in setting the relationship code to bad data, and sometimes the roster member's age also

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

was set to bad data. In general, no effort was made to try to match values of roster-derived household composition variables between pair members, since interviews of the different members of the same household could have taken place at different times. However, information from other pair members was sometimes used to change a relationship code from one value to another, instead of setting the relationship code to bad data. The editing of the household roster is described in detail in Chapter 8 of the imputation report (Aldworth et al., 2007).

6.2.2 Creation of the Pair Relationship Variable PAIRREL

Because the creation of the multiplicity factors was not automatic, multiplicities could not be created for all possible pair relationships. The following pair relationships were considered "of interest," requiring the creation of multiplicities in each case.

a. Parent-child, child aged 12 to 14

b. Parent-child, child aged 12 to 17

c. Parent-child, child aged 15 to 17

d. Parent-child, child aged 12 to 20

e. Sibling-sibling, younger aged 12 to 14, older aged 15 to 17

f. Sibling-sibling, younger aged 12 to 17, older aged 18 to 25

g. Spouse-spouse (includes partner-partner), with children younger than 18[10]

h. Spouse-spouse (includes partner-partner), with or without children[2]

Even though these pair relationships were of the most interest, all types of pairs were selected. The identification of the relationships was limited by the relationship codes that were available: parent, child, grandparent, grandchild, sibling, spouse, live-in partner, roommate, parent-in-law, child-in-law, boarder, other relative, and other nonrelative. This precluded the possibility of identifying an uncle-nephew relationship, for example. The various pair relationships that could be identified are given in the variable PAIRREL, the levels of which are summarized in Table 6.1. The levels in PAIRREL do not correspond exactly with those given above, but the relevant pair relationships can be derived from the value of PAIRREL. For example, a value of PAIRREL = 3 indicates that, among the pair relationships given above, the pair relationship was a parent-child pair with a child between 12 and 20 years old.

The process of identifying the pair relationships was a three-step process: (1) match the household rosters of the pair members, (2) determine the pair relationships using the relationship codes and ages of the matched rosters if they could be determined, and (3) impute missing pair relationships and create a final imputation-revised pair relationship variable. The first step is described in Section 6.2.2.1 and Appendix O, and the second is described in Section 6.2.2.2 and Appendix O. For the third step, covariates had to be created for the imputation models. The

[10] The spouse-spouse pair relationship included respondents who were legally married as well as respondents who lived together as though married ("partners"). Although the questionnaire distinguished between "spouses" and "partners," the pair relationship variable being described here did not distinguish between the two. In rare instances, a spouse-spouse pair included one pair member who identified the second pair member as a spouse, whereas the second pair member identified the first as a partner.

creation of these covariates is described in Section 6.2.3 and Appendix P, and the imputations are described in Section 6.2.4. Model summaries for the imputation models are provided in Appendix Q.

Table 6.1 Levels of the Variable PAIRREL

Value of PAIRREL	Interpretation	Domain of Interest?
1	The respondent is part of a parent-child (12–14) pair.	Yes
2	The respondent is part of a parent-child (15–17) pair.	Yes
3	The respondent is part of a parent-child (18–20) pair.	Yes, indirectly
4	The respondent is part of a parent-child (21+) pair.	No
5	The respondent is part of a sibling (12–14)-sibling (15–17) pair.	Yes
6	The respondent is part of a sibling (12–17)-sibling (18–25) pair.	Yes
7	The respondent is part of another sibling-sibling pair.	No
8	The respondent is part of a spouse-spouse[1] pair, with children in the household younger than the age of 18.	Yes
9	The respondent is part of a spouse-spouse[1] pair, with no children in the household younger than the age of 18.	Yes
10	The respondent is part of a spouse-spouse[1] pair, but it is unclear whether children younger than the age of 18 in the household belong to the pair.	Yes
11	The respondent is part of a grandparent-grandchild pair.	No
12	The respondent is part of another clearly identifiable pair.	No
13	The respondent is part of a pair that is not clearly identifiable, but it is clear from the relationship codes that it is not within codes 1 through 11.	No
14	The respondent is part of a pair that is not clearly identifiable, and it could be any pair relationship.	Maybe
15	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 1 or 12.	Maybe
16	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 2 or 12.	Maybe
17	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 3 or 12.	Maybe
18	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 4 or 12.	No
19	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 5 or 12.	Maybe
20	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 6 or 12.	Maybe
21	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 7 or 12.	No
22	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8 or 12.	Maybe
23	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 9 or 12.	Maybe

Table 6.1 Levels of the Variable PAIRREL (continued)

Value of PAIRREL	Interpretation	Domain of Interest?
24	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8, 9, or 12.	Maybe
25	The respondent is part of a pair that is not clearly identifiable, but it could be pair codes 11 or 12.	No
99	The respondent is not a member of a pair.	No

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

6.2.2.1 Matching the Household Rosters

To match the household rosters of the pair members, let the pair members be identified as pair member "A" and pair member "B." For the household roster of pair member A, it was necessary to determine which listed household member in A's roster corresponded to the other selected pair member. The same had to be done for pair member B. This was accomplished using the age and gender of the pair members, in addition to a variable (hereafter referred to as MBRSEL) that was supposed to identify the roster member corresponding to the other selected pair member. In a perfect setting, the questionnaire age and gender of pair member B (AGE and IRSEX, respectively) would have corresponded exactly to the age and gender entered for one of the members of pair member A's household roster (RAGE and RSEX). Moreover, the value of MBRSEL for this roster member would have been 1, and the value of MBRSEL for all other roster members would have been 0 or missing. In this perfect setting, exact matches with exactly one MBRSEL = 1 correctly identifying the other pair member also would have been found with pair member B's roster. This did not always occur, of course, so some effort was required to determine the roster member most likely to correspond to the other selected pair member.

In fact, the quality of the match varied depending upon the quality of the roster entries and the time between interviews. There are a number of if-then-else conditions, called priority conditions (due to the hierarchical nature of the conditions), each of which gave a pair match that was considered valid in the vast majority of cases. These conditions are provided in Appendix O. In general, the conditions matched IRSEX and AGE for the one pair member against the age and sex of the roster members in the other pair member's roster, using MBRSEL to help identify the appropriate roster member. These conditions in general terms are provided in Table 6.2. It was necessary that at least one of the two pair members have a match as good as that shown below.

Table 6.2 Measures of the Quality of Definitive Roster Matches

Measure Number	Description
0	Age and gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
1	Age and gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
2	Age within one, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
3	Age within two, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member
4	Age and gender matched exactly, with MBRSEL missing for all roster members
5	Age matched exactly, gender off, with exactly one MBRSEL correctly identifying the other pair member
6	Age within one, gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
7	Age within two, gender matched exactly, with MBRSEL correctly identifying the other pair member, but there was more than one MBRSEL[1]
8	Age within one, gender matched exactly, with MBRSEL missing for all roster members
9	Age within two, gender matched exactly, with MBRSEL missing for all roster members[2]
10	Age within 10, gender matched exactly, with exactly one MBRSEL correctly identifying the other pair member[2]

[1] 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. As a result, measures #1, #6, and #7 did not occur in the 2005 survey.
[2] For pairs where one pair member had a match corresponding to measures #9 or #10, if the other pair member had a match no better than measure #9, an additional requirement was implemented where the reported household sizes for both pair members had to be equal to 2.

Given that at least one side had a match according to one of the measures provided in Table 6.2, the other side could have a match that was weaker (i.e., not definitive), using the measures in Table 6.3. Additional columns are provided in Table 6.3, showing the weakest match that was allowed (as denoted by the measure) for the other pair member. The column titled "In Code" shows the weakest measure allowed in the code, and the column titled "Observed" shows the weakest measure that was actually observed for the other pair member.

In the cases where a single roster member had to be selected among duplicates (measures #14, #15, #17, and #18), where the duplicates had the same relationship code, it was necessary that the relationship codes be limited to child or sibling.

In some cases, due to the poor quality of the rosters of the pair members, it was not possible to locate the listed household member in A's roster that corresponded to pair member B, and vice versa. The determination of the pair relationships for these cases was left to imputation. Even when a pair of roster members was successfully identified, it was not always possible to successfully determine the pair relationship, as is pointed out in the next section.

Table 6.3 Measures of the Quality of Roster Matches That Are Not Definitive, Given That One Side Had a Definitive Match (as Shown by the Conditions Provided in Table 6.2)

Measure Number	Description	Weakest Measure Allowed for Other Pair Member	
		In Code	Observed
11	Age within 10, gender matched exactly, with MBRSEL missing for all roster members, provided another roster member with a closer age could not have been chosen	8	0
12	Everything missing, but the other pair member had good data	9	4
13	Age missing, gender matched exactly, household sizes equal	9	4
14	Age and gender matched exactly for two roster members, both with the same relationship code, with two MBRSELs identifying the two roster members (one was randomly selected)[1]	8	Not observed
15	Age and gender matched exactly for two roster members, both with the same relationship code, but MBRSEL was missing for all roster members (one was randomly selected)	8	0
16	Age, gender, and relationship code matched exactly for two or more roster members, and MBRSEL was missing for all roster members (one was randomly selected)	8	Not observed
17	Age within one and gender matched exactly for two roster members, both with the same relationship code, with two MBRSELs identifying the two roster members (one was randomly selected)[1]	8	Not observed
18	Age within one and gender matched exactly for two roster members, both with the same relationship code, with two MBRSELs identifying the two roster members (one was randomly selected)[1]	8	Not observed
19	Age within one, gender off, with exactly one MBRSEL correctly identifying the other pair member, and only two members in household	10	2
20	No matches possible, but relationship codes indicate the pair is not a part of a domain of interest	As with other pair member	As with other pair member
21	Age matches exactly, gender off, with MBRSEL missing for all roster members	9	0
22	No matches possible	9	4

[1] 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. As a result, measures #14, #17, and #18 did not occur in the 2005 survey.

6.2.2.2 Determining the Pair Relationship Using the Relationship Codes of the Matched Rosters

Once the pair was identified, two observations per household resulted, each with a relationship code corresponding to the other selected pair member. The relationship codes for these two observations had to be matched to determine the pair relationship. For example, suppose a 15-year-old and a 38-year-old were selected. If the 38-year-old was identified as the parent on the 15-year-old's roster, and the 15-year-old was identified as the child of the 38-year-old on the 38-year-old's roster, then the pair relationship would be identified as PAIRREL = 2 according to the levels of PAIRREL provided in Table 6.1. Thus, these two individuals would belong to the following pair relationships of interest: child (15 to 17)-parent, child (12 to 17)-parent, and child (12 to 20)-parent. As noted earlier, the pair relationship of interest was derived from the values of PAIRREL. In particular, the child (12 to 17)-parent and child (12 to 20)-parent domains were derived from pair relationships created using 12- to 14-year-olds, 15- to 17-year-olds, and 18- to 20-year-olds, the levels referenced in the levels of PAIRREL. Moreover, the overall spouse-spouse domain was derived from the two spouse-spouse pair relationships with and without children.[11]

As with the procedure used to match the household rosters, a series of if-then-else conditions were used to identify the relationship between pair members. These conditions, also called priority conditions because of their hierarchical nature, used ages and relationship codes to identify the pair relationships and are summarized in Appendix O. In a perfect setting, like the example given in the first paragraph of this section, the relationship codes would be nonmissing and in agreement between the pair members. In some instances, however, either the relationship codes were missing, or they did not agree across the pair members. The detailed conditions provided in Appendix O present a method for interpreting the relationship codes in such cases.

A few points that summarize the strategies used to identify a pair relationship in an imperfect setting follow:

1. If a relationship code was missing on one side but not on the other, the pair relationship was assumed to be identified by the nonmissing relationship code. The exception to this rule occurred if the identified relationship was child-parent with a child younger than 18, the "parent" was less than 10 years older than the child, and the "parent" answered the parenting experiences question (FIPE3) by saying that the other respondent was not his or her child. In this case, the nonmissing relationship code was considered spurious, and the relationship was left missing.

2. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but the relationship codes on both sides indicated that the unknown pair relationship was not a relationship of interest, the pair relationship was identified as such and no imputation was required. For example, if pair member "A" identified pair member "B" as a "boarder," but pair member "B" identified pair member "A" as "other relative," the relationship was not a relationship of interest, hence code "13" would have been applied in the variable PAIRREL.

[11] The spouse-spouse pair relationship includes partner-partner pair relationships.

3. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but a parent-child relationship was possible given the relationship code in one of the pair member's rosters, the FIPE3 variables were used to assist in the determination of a pair relationship. An example of a case where this would have been useful is a pair member who was a stepparent refers to his or her stepchild as "child," but the child refers to the stepparent as "other nonrelative." Membership in a parent-child relationship where the child was younger than 18 was indicated if the stepparent answered FIPE3 affirmatively, thereby entering the parenting experiences module. On the other hand, if the stepparent answered FIPE3 negatively, then the stepparent was not considered the parent. A third scenario arose if the FIPE3 answer was not given. In this case, a parent-child relationship was assumed if the stepparent was legally married and the child identified the spouse of the other pair member as "parent."

The quality of the match for PAIRREL levels 1 through 25 is indicated by the variable RELMATCH, the levels of which are summarized in Table 6.4.

In general, imputation was required for values of RELMATCH of 0 or 4 or if PAIRREL = 10. PAIRREL = 10 was a special case, since it was clear that a relationship "of interest" always would have been involved. For this value of PAIRREL, the value of RELMATCH was equal to 1 or 2. However, imputation was still required since it was not clear whether children were in the household. The number of cases that were matched or not matched, as indicated by the RELMATCH variable (or PAIRREL = 10), is provided in Table 6.5 for the 2005 survey. The amount of imputation required was dependent upon the quality of the rosters. The attributes of the roster are described in Chapter 8 of the imputation report (Aldworth et al., 2007).

Table 6.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH

Value of RELMATCH	Values of PAIRREL	Interpretation
0	**14**	FAILURE: The relationship was not identifiable and could have been a relationship of interest.
1	**1–9, 11–13**	SUCCESS: The relationship was clearly identifiable using information from both pair members or was unmistakably not a relationship of interest.
1	**10**	FAILURE: A spouse-spouse[1] relationship was definitively established using information from both pair members, but it was unclear whether the pair had children in the household.
1.5	**8**	SUCCESS: A spouse-spouse[1] relationship was definitively established using information from both pair members, and children younger than 18 were in both rosters. Relationship codes on one side indicated children belonged to the pair, and on the other side, the relationship codes corresponding to the children were missing.
2	**1–9, 11–13**	SUCCESS: The relationship was clearly identifiable using information from one pair member, while the relationship code from the other pair member was missing.

Table 6.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH (continued)

Value of RELMATCH	Values of PAIRREL	Interpretation
2	10	FAILURE: A spouse-spouse[1] relationship was definitively established using information from one pair member, while the relationship code from the other pair member was missing. It was unclear whether the pair had children in the household.
3	1–7, 8, 12, 13	SUCCESS: Relationship information was conflicting between the pair members, but conclusions were drawn anyway for some parent-child pairs, some sibling-sibling pairs, and some spouse-spouse[1] pairs using either information outside the household roster or logical reasoning.[2]
4	15–25	FAILURE: Relationship information was not identifiable. Information was in conflict between the pair members, where one pair member indicated relationship of interest and the other did not. However, ages supported a relationship of interest (may be used to limit imputation).

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.
[2] In the case of potential parent-child pairs, further evidence that a parent-child relationship was involved or not involved was obtained by looking at the FIPE3 variable, by whether a stepparent had a spouse that corresponded to a child's parent or by the ages of the respondents. For spouse-spouse relationships, two situations occurred: in the case where the respondents were not legally married, the children of one pair member were considered the children of the pair in the household, even though they were not identified as such by the other pair member. In the case where only one pair member referred to the other as a "married" or "unmarried partner," if both had the same children, they were considered "spouse-spouse-with-children." The other pair member was usually referred to as a "roommate" or "other nonrelative."

Table 6.5 Frequencies of the Levels of the Variable RELMATCH: 2005

RELMATCH	2005
0	28 (0.14%)
1 (PAIRREL ≠ 10)	19,996 (97.95%)
1 (PAIRREL = 10)	32 (0.16%)
1.5	0 (0.00%)
2 (PAIRREL ≠ 10)	69 (0.34%)
2 (PAIRREL = 10)	0 (0.00%)
3	122 (0.60%)
4	168 (0.82%)

6.2.3 Creation of Covariates for Imputing Pair-Level Variables

For pairs where the relationship was not clear due to missing pieces of the household roster, or where pairs could not be determined because the relationship codes did not match, imputation was required. In stages two and three, imputation also was required for missing multiplicities and household-level person counts. In all three stages, the PMN method was used to impute missing values, which required the fitting of models. Since the imputation was performed at the household level rather than at the respondent level, it was necessary to have classing variables (i.e., variables forming imputation classes) and model covariates defined at the household level. Segment-level covariates were used for this purpose since they were automatically defined at the household level, using external information that was constant regardless of when the interviews were conducted. However, information from the questionnaire also would have been useful. Logical choices for questionnaire-derived variables would be the household composition variables IRHHSIZE (household size), IRKID17 (number in household younger than the age of 18), IRHH65 (number in household aged 65 years or older), and IRFAMSKP (indicator whether other family members in household). However, because interviews between pair members could have been conducted at different times, these variables were not necessarily consistent across pair members. New count variables were needed that were consistent across the pair members within a household, which used the screener information to reconcile disagreements between pair members. These variables were created in two steps: (1) create the count variables for each pair member, and (2) attempt to reconcile disagreeing values between pair members. The following sections describe these two steps in the creation of household size, household composition age count variables, and household composition age count variables for males only, each of which were consistent across pair members. These variables also had to be created for respondents who were not part of a pair, for the purposes of creating and imputing the household-consistent person counts of various domains.

6.2.3.1 Household Size

The new variable created to represent a household size that was consistent across the pair members was called HHSIZE. The first step was to compare the edited household size, TOTPEOP, between pair members. If the values for TOTPEOP agreed across pair members and were both nonmissing and greater than 1, then HHSIZE was simply set to that value. There were two ways that TOTPEOP would disagree across pair members. In the first case, if the count for one pair member was missing, and the count for the other was not and was greater than 1, a natural choice for HHSIZE would have been the nonmissing value. In the second case, the household size counts disagreed across pair members. The tools used to determine the final value of HHSIZE in these cases included the reported and edited household size variables previously mentioned, as well as other measures of household size and "quality of roster" measures. These "other measures" included the screener household size and two sums of total valid ages within a pair member's roster. The first sum was a simple total count of the number of roster members with valid ages, obtained by summing the counts within certain age groups. The second sum adjusted the first by accounting for the minimum number within each age category, given the questionnaire ages of the two pair members. It differed from the first if the number of valid ages in a given age category was less than the minimum possible in that age category, given the ages of the two pair members selected. For example, suppose a household roster had one 12- to 17-year-old, but two 12- to 17-year-olds were selected. The second sum was determined by replacing the number of 12- to 17-year-olds by the minimum number possible, 2. An additional situation occurred where the household size counts could not be easily determined by looking at both pair members. If the counts for both pair members were missing, the screener household size was used to define HHSIZE. In some cases, disagreement between pair members with regard to the true household size could not be easily resolved. The screener household size did not support either household size in these cases, and the age counts mentioned above also did not resolve the disagreement. A decision had to be made as to which pair member's household size should be believed. This decision depended upon the "quality of the roster," where the household size was determined by the pair member with a better "roster quality." One obvious way to measure roster quality was by noting the number of cases where the ages, relationship codes, or genders were missing in the roster. Clearly, if a roster was missing one or more of these three variables for some of the roster members, the roster was of "poorer quality" than a roster with these variables nonmissing for all roster members.

If only one household member was selected as a respondent, known colloquially as a "nonpair household," the rules for creating HHSIZE were the same as those that were used if two household members were selected in a pair, but only one of the pair members had a nonmissing, acceptable value for a reported household size, with one important exception. If only one household member was selected as a respondent, it was obviously permissible to have a reported household size of 1, whereas in a selected pair a reported household size of 1 was considered "bad data," necessitating the use of the screener household size as the source variable for HHSIZE.

In summary, the variables used to determine HHSIZE included, for each pair member, the reported and edited household sizes, the number of cases with valid ages in the roster, the number of cases with valid ages with the count in some age categories replaced by the minimum possible in that age category, and a quality of roster count of the number of roster members with missing information. The screener household size, which was the same for each pair member, also was used. Using all of these tools, HHSIZE did not have any missing values in the 2005 survey, nor did it have any in surveys from previous years. General points about the creation of the household size variable are provided in Appendix P.

6.2.3.2 Household Composition Age Count Variables

It would seem logical to assert that the ages of other household members would be good predictors for the domain to which a pair might belong. Such variables also would be important for imputing multiplicity and household-level domain counts. The household-consistent age counts were limited to the following age ranges: younger than 12 years old, 12 to 14 years old, 15 to 17 years old, 12 to 17 years old, 12 to 20 years old, 18 to 25 years old, 26 to 34 years old, 35 to 49 years old, and 50 years old or older. These variables were called AGE011, AGE1214, AGE1517, AGE1217, AGE1220, AGE1825, AGE2634, AGE3549, and AGE50P, respectively.

The first step in this process was to count the nonmissing ages for roster members in the household for each pair member. In some cases, it was necessary to adjust the count since the ages could not be matched exactly. For example, suppose a 38-year-old and a 17-year-old were interviewed and the 17-year-old was interviewed first. Suppose also that the 17-year-old turned 18 (i.e., had his 18^{th} birthday) before the 38-year-old was interviewed. Hence, the 17-year-old would have had an age of 18 in the 38-year-old's roster. Because the ages for the pair domains were defined at the time of each pair member's interview, the ages of interest for pair domains would have been 17 and 38. Hence, it was necessary to account for this by creating a new roster age variable that matched the age provided in the other pair member's questionnaire. The age counts using this new roster age variable were equivalent to subtracting 1 from the previously obtained 18- to-25 count and adding 1 to the previously obtained 12-to-17 count in the 38-year-old's roster. These adjustments were made for all cases where a match was made between one pair member's roster and another pair member's interview age and sex and the ages did not match exactly.

If no roster ages were missing, the sum of these counts was equal to the edited household size, TOTPEOP. Note that the raw household size was not considered here, since the counts were obtained from an edited roster. As with household size, a series of if-then-else conditions were used to obtain the most likely count within each age group. These conditions are called priority conditions due to their hierarchical nature. If the appropriate count was ambiguous due to disagreement between the pair members, the quality of the roster and the age of the respondent (in that order) were used to determine the appropriate count. The roster quality was determined by the number of bad or missing roster entries (as indicated in the previous section) and the quality of the match between the pair member's roster and the other pair member's questionnaire age and sex. If only one household member was selected as a respondent, the rules were the same as when two household members were selected in a pair, but only one of the pair members had nonmissing data for the roster ages, with one important exception. When determining minimum possible counts for various age groups, it was obviously not necessary to incorporate information from another pair member to increment the minimum for that pair member. General points about the creation of the age variables are provided in Appendix P.

6.2.3.3 Household Composition Age Count Variables for Males Only

For some pair variables, particularly spouse-spouse pairs, knowledge of the gender of the roster member was important in imputing missing values. In a similar manner to that used in the creation of the household composition age count variables, variables counting the number of males within the given age ranges were created. Disagreements between pair members were resolved in a similar manner to what was done with the household composition age count variables, as described in the previous section. For a given age range, the number of females could be obtained by subtracting the number of males from the total number within that age range. The names of the male age counts were MALE011, MALE1214, MALE1517, MALE1217, MALE1220, MALE1825, MALE2634, MALE3549, and MALE50P.

6.2.4 Creation of the Imputation-Revised Pair Relationship Variable IRPRREL

It was not always possible to definitively determine the pair relationship for the selected pair. In some cases, the relationship codes between the two pair members could not be reconciled. In other cases, no information was available at all about the type of pair relationship. This section describes how those missing pair relationships were imputed using the PMN method described in Appendix N. In this section, the application of the PMN method to the imputation of pair relationships is described. Since only the pair relationship was imputed, the imputation was univariate in the sense that no sequential models were necessary. However, in some cases the outcome variable was multinomial, which meant that matching was done on more than one predicted mean for each recipient pair. The name given to the imputation-revised pair relationship variable was IRPRREL.

6.2.4.1 Setup for Model Building

Pair relationships varied greatly according to the age of the respondent. Table 6.6 presents 11 age group pairs, followed by the pair relationships prevalent within each age group pair. The widely varying distributions of pair relationships within each age group pair are evident in this table. Because of the different prevalence of pair relationships within age group pairs, PMN was applied separately within each age group pair. Imputations were done one variable at a time so that no hierarchy of variables was required to set up a sequence of models, as is normally done with PMN. The first step, therefore, was to define respondents, nonrespondents, and the item response mechanism within each age group pair. For a pair to be considered a complete data responding pair, the pair relationship must be definitively established. In terms of the variable PAIRREL, this meant that the pair had to have a value of PAIRREL within the range of 1 to 9, or equal to 11 or 12. A value of PAIRREL equal to 13 also was considered complete, even though the pair relationship was not definitively established, since it was known that the pair relationship was not a relationship of interest. Response propensity adjustments then were computed for each age group pair in order to make the respondent pair weights representative of the entire sample of pairs. (Because the modeling of the final pair weight adjustments was not completed at the time of the pair imputations, the pair-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.)[12] These adjustments were calculated using an item response propensity model. This model is a special case of GEM, which is described in greater detail in Appendix A.

Table 6.6 Age Group Pairs with Associated Possible Pair Relationships

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1]	
		≥ 10% Prevalence[2]	< 10% Prevalence[3]
0	12–14/12–14	Sibling-sibling	Other relationship
1	12–14/15–17	Sibling-sibling	Other relationship
2	12–14/18–25	Sibling-sibling	Other relationship; parent-child; spouse-spouse**
3	15–17/15–17	Sibling-sibling	Other relationship; spouse-spouse*
4	15–17/18–25	Sibling-sibling	Other relationship; spouse-spouse; parent-child*
5	18–20/18–25	Other relationship; sibling-sibling; spouse-spouse	Parent-child**
6	21–25/21–25	Spouse-spouse; other relationship; sibling-sibling	Parent-child**
7	12–14/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling*
8	15–17/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling; spouse-spouse**

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

Table 6.6 Age Group Pairs with Associated Possible Pair Relationships (continued)

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1]	
		≥ 10% Prevalence[2]	< 10% Prevalence[3]
9	18–20/26+	Parent-child	Other relationship; sibling-sibling; spouse-spouse; grandparent-grandchild
10	21+/26+	Spouse-spouse; parent-child; other relationship; sibling-sibling	Grandparent-grandchild*

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships. The spouse-spouse domain as listed here actually consists of two domains (spouse-spouse-with-children and spouse-spouse-without-children) that have been collapsed for the purposes of making the table easier to read. "Other relationship" refers to a relationship other than sibling-sibling, parent-child, grandparent-grandchild, or spouse-spouse.
[2] The pair relationships in this column each form at least 10 percent of the overall total number of pair relationships with the given age group pair, and the total is at least 85 percent of the overall total.
[3] Pair relationships followed by stars occur rarely, in less than 1 percent of the overall total number of pair relationships. Two stars indicate such rarity that the pair relationship did not appear in the age group pair in every survey year.

6.2.4.2 Model Building and Determination of Predicted Means

The PMN method is a two-step process. The first step is the modeling step, followed by a hot-deck step where imputed values replace missing relationships. As stated earlier, each age group pair acted as an imputation class, within which the modeling and hot-deck steps were performed separately. The different attributes of the 11 models, corresponding to the 11 age group pairs, are described in this subsection.

Response categories. Ideally, each type of pair relationship within an age group pair would constitute a response category in a multinomial response model. However, the number of cases corresponding to some pair relationships within each age group pair were very small, as is apparent in Table 6.6. Hence, it was not feasible to fit multinomial models that cover all the possible pair relationships for a given age group pair. Rather, in the modeling step, some of the response categories were combined with separate assignments of imputed values within each of the 11 age group pairs. Priority was placed on placing the pair relationships "of interest" into separate categories. In some cases, pair relationships that were not of interest were combined with other categories, even if there were sufficient numbers to have a separate category in the multinomial model. Table 6.7 presents the response categories that were used for modeling. The delineation between categories that were combined for modeling was left to the hot-deck step.

As an example, consider age group pair #5. In this age group pair, there are typically four types of pair relationships that have a sufficient number of respondent pairs to fit a satisfactory model, including spouse-spouse domains, sibling-sibling pairs, and all others. However, it is always easier to fit a good model with a smaller number of levels in the response. Since only two of those four were pair relationships of interest, these two (the two spouse-spouse domains) were used as levels in the response variable. The third level was obtained by combining the sibling-sibling and other relationship pairs. There are typically a small number of parent-child pairs, which also were combined with the other relationship pairs.

Table 6.7 Modeled Pair Relationships within Age Group Pairs

Age Group Pair Number	Age Group Pair	Number of Levels in Response	Levels of Modeled Response
0	12–14/12–14	2	Sibling-sibling; all others
1	12–14/15–17	2	Sibling-sibling; all others
2	12–14/18–25	2	Sibling-sibling; all others
3	15–17/15–17	2	Sibling-sibling; all others
4	15–17/18–25	2	Sibling-sibling; all others
5	18–20/18–25	3	Both spouse-spouse pair relationships;[1] all others
6	21–25/21–25	3	Both spouse-spouse pair relationships;[1] all others
7	12–14/26+	2	Parent-child; all others
8	15–17/26+	2	Parent-child; all others
9	18–20/26+	2	Parent-child; all others
10	21+/26+	3	Both spouse-spouse pair relationships;[1] all others

[1] The two spouse-spouse pair relationships are the spouse-spouse and the spouse-spouse-with-children–younger-than-18 pair relationships. The pair relationships labeled "spouse-spouse" include partner-partner pair relationships.

Covariates in models. After the weights were adjusted using the item response propensity model within each age group pair, binomial and multinomial logistic models were fitted using the adjusted weights with the response variable defined as in Table 6.7. As noted in previous sections, the number of covariates at the household level was limited. The pool of covariates used in the item response propensity model included the following variables:

1. household size (HHSIZE, as defined in Section 6.2.3.1),
2. age category of older respondent (where applicable),
3. race of older respondent,[13]
4. sex of older respondent,
5. sex of younger respondent,
6. marital status of older respondent (where applicable),[14]
7. marital status of younger respondent (where applicable),
8. education of older respondent (where applicable),[15]
9. education of younger respondent (where applicable),
10. employment status of older respondent (where applicable),[16]
11. employment status of younger respondent (where applicable),
12. region,

[13] Race had four levels: white, black/African American, American Indian/Alaska Native, and Asian/Pacific Islander.

[14] Marital status had four levels for respondents 15 or older: married, widowed, divorced, and never married.

[15] Education had four levels for respondents 18 or older: less than high school, high school graduate, some college, and college graduate.

[16] Employment status had four levels for respondents 18 or older: full-time employed, part-time employed, unemployed, and other (not in work force).

13. MSA (metropolitan statistical area),
14. categorical percent Hispanic or Latino in segment,
15. categorical percent black or African American in segment, and
16. categorical percent owner-occupied households in segment.

In some cases, due to the ages of the pair members, the education, employment status, and marital status did not apply to one or both members of a pair. In order to increase the ability to obtain convergent models, some of the cells in the categorical covariates were collapsed.

Additional variables defined in Section 6.2.3.2 were used to adjust the weights in the final response models for each of the 11 age group pairs in those cases where the variables were nonmissing. The variables follow:

1. number in household aged 0 to 11,
2. number in household aged 12 to 17,
3. number in household aged 18 to 25,
4. number in household aged 26 to 34,
5. number in household aged 35 to 49, and
6. number in household aged 50+.

In the cases where these variables were all nonmissing, they were put into the pool of covariates for the final response model in place of HHSIZE. However, there were a handful of cases for which these variables could not be determined. In those cases, 11 additional final response models were fitted without the household composition age count variables listed above, using the same pool of covariates that were used for the item response propensity models.

Building of models. For age group pairs 0 through 4 and 7 through 9, binary logistic regression models were built. Since there were three outcomes with age group pairs 5, 6, and 10, multinomial polytomous logistic models were fitted for these age group pairs. All the models incorporated the design pair weights that were ratio adjusted for unit nonresponse (where a pair was selected but did not respond to the survey) and calibrated to account for item nonresponse (where a pair responded to the survey, but the pair relationship was unknown), using the item response propensity models, as described in Section 6.2.4.1. Naturally, not all of the covariates in the original pool could be included in each model due to convergence problems. The final set of covariates corresponding to each model is provided in Appendix Q.

Determination of predicted means. Although models were built using respondent pairs where the pair relationship was known definitively, predicted probabilities were required for all pairs. Once the models were fitted, predicted means were determined for both respondent pairs and nonrespondent pairs, using the parameter estimates from the models.

6.2.4.3 Constraints on Hot-Deck Neighborhoods and Assignment of Imputed Values

If possible, donor pairs in the hot-deck step of PMN were chosen with predicted means within delta[17] of the recipient pair's predicted mean(s), where the value(s) of delta varied depending upon the value of the predicted means. In this case, delta was defined as 5 percent of the predicted probability if the probability was less than 0.5, and 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5 and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is shown in Table 6.8. If no donor pairs were available with predicted means within delta of the recipient pair's predicted mean, the neighborhood was abandoned, and the donor pair with the closest predicted mean was chosen.

Table 6.8 Values of Delta for Various Predicted Probabilities

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

In general, the members of the neighborhoods were restricted to satisfy two types of constraints: "logical constraints" and "likeness constraints." Constraints that made the imputed values consistent with preexisting values of other variables were called logical constraints and were required for the candidate donor pair to be a member of the neighborhood. Likeness constraints were implemented to make donor pairs and recipient pairs as much alike as possible. Although logical constraints could not have been loosened, likeness constraints could have been loosened if they had forced the donor pool to be too sparse. Details of these imputation procedures are provided in Appendix N.

In addition to the likeness constraint defined by delta, other likeness constraints also were included in the neighborhoods. These constraints follow:

Older pair member age constraint, 26+-year-old pair members. The 26+ age group, associated with age group pairs 7 through 10, was split up into three groups: 26 to 34, 35 to 49, and 50+. This was most useful to delineate child-parent pairs.

Marital status likeness constraints. Each respondent's marital status, as entered in the core section of the questionnaire, was closely related to the relationship between the pair members.[18] This marital status variable had four levels among respondents aged 15 or older: married, widowed, separated or divorced, and never married. Marital status likeness constraints combined the information from this variable for both pair members, where the levels were

[17] "Delta" refers to the value that defined the neighborhood of donor pairs that were "close" to the recipient pair. The difference between the predicted mean of the recipient pair and the predicted means of the donor pairs must have been less than delta. See Appendix N for more details.

[18] Pairs that included a pair member with an imputed marital status were not eligible to be donor pairs. If a recipient pair had a pair member with an imputed marital status, then donor pairs had any marital status, unless one of the pair members in the recipient pair had a nonimputed marital status indicating married, widowed, or divorced.

collapsed in different ways depending upon the age group pair. For age group pairs where the only pair relationships of interest involved were child-parent pairs, two classes were required: both respondents never married and one respondent never married. Three classes were required for age group pairs where the only pair relationships of interest involved were spouse-spouse pairs: both respondents not currently married, one respondent not currently married, and both respondents currently married. Finally, six classes were attempted if both the spouse-spouse and child-parent pair relationships were possible: (1) both respondents never married; (2) one respondent never married, the other formerly married (widowed or divorced); (3) one respondent never married, the other currently married; (4) both respondents formerly married (widowed or divorced); (5) one respondent formerly married, the other currently married; and (6) both respondents currently married. It should be noted that not all of these classes would need donor pairs if no recipient pairs were within the class. It also should be noted that marital status could not have been considered a logical constraint where spouse-spouse pairs were involved, since many live-in partners (who were considered spouse-spouse pairs) answered the marital status question as "never married."

Gender makeup of pair likeness constraints. For donors who formed a spouse-spouse pair, the vast majority were male-female. Hence, in those cases where a spouse-spouse pair was possible, the gender likeness constraint required that the donor pair and recipient pair be either both of the same gender or both of a different gender. This meant that the likelihood of same-sex spouse-spouse pair relationships were equally likely (more or less, depending upon the model) among donors and recipients.

Age constraints on 15- to 17-year-old pair members. For the 15-to-17 age group, the likelihood of being in a spouse-spouse relationship was very small. Nevertheless, the likelihood that a 17-year-old was married was considerably greater than the likelihood for a 15-year-old. Hence, for the age group pairs where at least one pair member was between 15 and 17, the younger pair member of both the donor pair and recipient pair had to be of the exact same age.

Constraints on number of children. In Section 6.2.3.2, a covariate was defined for the number of children in the household younger than 12, AGE011, and one was defined for children in the household between 12 and 17, AGE1217. If there was disagreement between pair members on the values of these covariates, the pair member with information agreeing with the screener was used if possible. For the imputation of spouse-spouse relationships with and without children, these covariates were used to restrict donor pairs, where AGE011 was used for potential parents younger than 18, and AGE011+AGE1217 was used for potential parents 18 or older. If the recipient pair had no children according to the relevant covariate or covariates, donor pairs also did not have children. If the recipient pair had children, the same was true for the donor pair. In almost all cases, when there was disagreement between pair members regarding whether the pair had children in the household or not, the imputation used information that was closer to the screener.[19]

The likeness constraints were loosened in the following order (where applicable): (1) for the age group pairs where six marital status classes were used, collapse to two classes (the same

[19] This will not always be true, because it is not always possible that the screener can be used to determine the value for AGE011 and AGE1217 when the pair members' information disagrees.

two used when the only pair relationships of interest were child-parent pairs) or three classes (the same two used when the only pair relationships of interest were spouse-spouse pairs), depending on the response that was most common; (2) abandon the neighborhood and choose the donor pair with the closest predicted mean or means; (3) loosen age constraint (26+ groups); (4) loosen the remaining marital status restrictions; and (5) simultaneously loosen the age constraints on 15- to 17-year-old pair members and the gender makeup likeness constraints. The constraint on the number of children in the household was never loosened. For the multinomial logistic models, a Mahalanobis distance was used to define the distance across the multiple predicted probabilities.

Logical constraints were limited to the information that was already known about the pair, as denoted by the level of the variable PAIRREL. If, for example, PAIRREL = 14, then no information was available about the identity of the pair relationship and no logical constraint was needed. On the other hand, if PAIRREL = 15, this meant that the pair relationship was either a child-parent pair where the child was aged 12 to 14, or it was some relationship other than spouse-spouse, parent-child, grandparent-grandchild, or sibling-sibling. One could argue that the household composition age counts be considered logical constraints. However, these variables did not exist for all respondent pairs, and in some cases the values were set for these variables in a somewhat arbitrary manner. Moreover, due to the timing of the interviews, it was conceivable that an unexpected pair relationship could occur even though the household composition age counts would seem to preclude it.

6.2.4.4 Additions to the Analytic File

The imputation indicator variable that accompanied IRPRREL was called IIPRREL, which summarized how the data in IRPRREL were obtained. In addition to these variables, the edited pair relationship variable PAIRREL, the quality-of-match indicator RELMATCH, and the pair indicator PAIRMEM, which simply indicated whether a respondent in the analytic file was part of a responding pair, were released to the analytic file. Four additional variables were released to the analytic file to aid in pair analyses. These included the variables PRNTIND, AGEOTHER, SEXOTHER, and PAIRID. PRNTIND identified whether the respondent was a parent in a parent-child relationship, AGEOTHER gave the age of the other respondent in the pair, SEXOTHER gave the gender of the other respondent in the pair, and PAIRID gave the questionnaire ID (QUESTID) of the other pair member.

6.3 Stage Two: Creation and Imputation of Multiplicities

As stated earlier, multiplicities were required to account for analyses that were made at the person level, even though the pair weights were calculated at the pair level. The multiplicities were relevant only at the person level, so naturally the definition of multiplicity required the identification of the focus member of the pair. Using the pair relationships determined in Section 6.2, the following domains were considered:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 15 to 17), parent focus;
4. parent-child (child 15 to 17), child focus;
5. parent-child (child 12 to 17), parent focus;

6. parent-child (child 12 to 17), child focus;
7. parent-child (child 12 to 20), parent focus;
8. parent-child (child 12 to 20), child focus;
9. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
10. sibling (12 to 14)-sibling (15 to 17), sibling (12 to 14) focus;
11. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
12. sibling (12 to 17)-sibling (18 to 25), sibling (12 to 17) focus;
13. spouse-spouse (includes partner-partner) with children younger than 18; and
14. spouse-spouse (includes partner-partner).

Determining the multiplicity entailed finding the number of roster pairs in the domain of interest that contained the focus member in the pair. In broad terms, the process of determining the multiplicity count was a three-step process: (1) determine the multiplicity count for each pair member; (2) use the screener, quality of roster, and other means to figure out the appropriate count if each pair member's counts did not match; and (3) impute multiplicities that otherwise could not be determined. The first step is described in Section 6.3.1, the second in Section 6.3.2 and Appendix R, and the third in Section 6.3.3. Model summaries for the imputation models are provided in Appendix Q.

Since the pair weights reflected selection done at the time of screening, the multiplicity count should have reflected the household makeup at that time. However, this was not entirely possible, since the screener roster was not as complete as the questionnaire roster, and recorded relationships in the screener roster were relative to the head of the household rather than to each pair member. Hence, no account was made for cases where a change in the household makeup occurred between the screening time and the time of both interviews. The change in household makeup could have occurred because of an intervening birthday or because a roster member left or entered the household after screening. Technically, adjustments should have been made to account for this. However, the number of cases where this occurred was small, and to implement such an adjustment would have been extremely complicated, especially for the household counts discussed in Section 6.4. Nevertheless, in cases where there were disagreements between pair members on the value of the multiplicity count, the screener was used to resolve those disagreements when possible.

6.3.1 Determining the Multiplicity Count for Each Pair Member

The multiplicity counts for each pair member consisted of a direct count and an indirect count. The direct count was obtained by looking at the pair member who was the focus. It was simply a count of the roster members that could have been selected, where the same pair domain would have resulted. The indirect count was obtained by looking at the pair member who was not the focus. It was a count of the pair member himself or herself, plus other roster members who, by virtue of their relationship code, would have had the same pair relationship had they been selected. A summary of the ways of determining the direct count and indirect count for each pair domain are provided in Table 6.9. For the domains provided in Table 6.9, neither the direct nor the indirect count could be 0, since the pair member who was not the focus had to be part of the count. For spouse-spouse counts, no work was necessary to determine multiplicity counts. If a respondent was in a spouse-spouse pair, the multiplicity count was necessarily 1 in almost all cases since only one spouse-spouse pair could have been selected that included that pair member. If the true multiplicity count exceeded 1, then the multiplicity count was set to 1.[20] Note that other spouse-spouse pairs in the household (one spouse's parents, for example) would have been of interest in the household counts discussed in subsequent sections.

Table 6.9 Multiplicity Counts for Each Pair Member

Pair Relationship	Focus Member	Direct Count	Indirect Count
Parent-Child	Child	From child: number of parents	From parent: self + spouse/partner
Parent-Child	Parent	From parent: number of children in appropriate age range	From child: self + number of siblings in the appropriate age range
Sibling-Sibling	Older sibling	From older sibling: number of siblings in younger age range	From younger sibling: self + number of siblings in younger age range
Sibling-Sibling	Younger sibling	From younger sibling: number of siblings in older age range	From older sibling: self + number of siblings in older age range

6.3.2 Determining the Final Multiplicity Count

Once the counts were determined for each pair member, it was necessary to resolve differences between these counts across pair members. In most cases, the direct and indirect counts agreed, with no bad relationship codes for either pair member, resulting in an easy determination of the final multiplicity count. An easy determination was usually possible if one pair member had bad relationship codes or had a count of 0, which meant that the final multiplicity count came from the pair member with good data.[21] Exceptions to this rule are described in Appendix R. For some cases, both pair members had bad relationship codes, which meant that the final multiplicity was left to imputation. Some of the remainder of cases could be reconciled and some could not. In the cases where reconciliation was possible, many of the disagreements between the pair members were resolved by going to the screener. The method used to reconcile differing counts depended upon the domain. In addition to the screener, for the parent-child domains, the FIPE3 variable was used to help reconcile differences. Detailed rules for reconciling differences between pair members are provided in Appendix R.

If reconciliation between the counts from the two pair members in the household and the screener was not possible, upper and lower bounds within which the imputed value had to reside were determined from the counts for each pair member and the counts for the screener. The amount of imputation required for the multiplicity counts is shown in Table 6.10 for the 2005 survey year. From this table, it is apparent that the greatest degree of uncertainty came with the determination of the number of parents in the child-focus parent-child domains. This occurred because, even though the parent-child pair relationship had been established, it often was unclear

[20] In rare cases, it was possible for a respondent to have two or more spouses. Determining the appropriate multiplicity count in these cases required knowledge of which spouse was the focus, which would be arbitrary. Because having multiple spouses was an extremely rare occurrence, and because of the complexity of determining the appropriate multiplicity count, these situations were not accounted for.

[21] There were some provisions to this rule. If the bad relationship codes were only within the relevant age ranges, then the count from the good side was used only if the age ranges in the good side matched the screener.

whether there was a second "parent" in the household. Other domains had very little uncertainty. The counts of the number of children in the parent-focus parent-child domain were almost always definitively determined.

Table 6.10 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2005

Pair Domain	Multiplicity	Missing Cases
		2005
Parent-Child (12-14), Child Focus	Number of parents	74
Parent-Child (12-14), Parent Focus	Number of children	0
Parent-Child (15-17), Child Focus	Number of parents	72
Parent-Child (15-17), Parent Focus	Number of children	0
Parent-Child (12-17), Child Focus	Number of parents	146
Parent-Child (12-17), Parent Focus	Number of children	0
Parent-Child (12-20), Child Focus	Number of parents	170
Parent-Child (12-20), Parent Focus	Number of children	0
Sibling (12-14)-Sibling (15-17), Older Sibling Focus	Number of younger siblings	2
Sibling (12-14)-Sibling (15-17), Younger Sibling Focus	Number of older siblings	2
Sibling (12-17)-Sibling (18-25), Older Sibling Focus	Number of younger siblings	6
Sibling (12-17)-Sibling (18-25), Younger Sibling Focus	Number of older siblings	2

6.3.3 Creation of Imputation-Revised Multiplicity Variables

In many cases where the pair relationships were not defined, multiplicity counts also were not defined. In addition, there were a handful of cases where multiplicity counts were not determined even when the pair relationship was known. In all of these cases, imputation was required to determine the multiplicity count. As with the pair relationship imputation, missing multiplicities were imputed using the PMN method described in Appendix N. In this section, the application of PMN to the imputation of multiplicities is described. Since only the multiplicity in the second stage was imputed for each pair, the imputation was univariate in the sense that no sequential models were necessary. However, in some cases, several variables were associated with a single model, as described below.

6.3.3.1 Setup for Model Building

Multiplicity counts were defined only within the relevant domain, which, in turn, depended upon the pair relationship. For the sibling-sibling pairs, four separate imputations were conducted for the multiplicities associated with the four sibling-sibling pair domains. The parent-child domains were hierarchical, however, where the imputations could not have been conducted independently if consistency was to be maintained. Hence, only two models were fitted to the child-parent pairs, using just the domains with children 12 to 20 years old. One set of models was for the number of the parent's children, and the other set was for the number of parents of the child. Using the predicted means from these models, a single donor pair for each focus was selected from which the multiplicity counts were determined for 12-to-14, 12-to-17, 15-to-17, and 12-to-20 child-parent pairs. No imputation was required for the spouse-spouse multiplicity counts, since a selected respondent in a spouse-spouse pair naturally had only one spouse.[22]

The first step for these six models was to define respondents, nonrespondents, and the item response mechanism for each model, separately. For a pair to be considered a complete data responding pair with regard to multiplicities, the multiplicity had to be nonmissing for all of the variables being imputed. For the parent-child pairs, this meant that the multiplicity had to be nonmissing for the domains with 12- to 20-year-olds. A nonmissing multiplicity for this domain would automatically guarantee nonmissing multiplicities for the subset parent-child domains. Response propensity adjustments were then computed for each of the six models in order to make the respondent pair weights representative of the entire sample of pairs. (Because the modeling of the final pair weight adjustments was not completed at the time of the pair imputations, the pair-level design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment.) These adjustments were calculated using an item response propensity model. This model is a special case of GEM, which is described in greater detail in Appendix A.

6.3.3.2 Model Building and Determination of Predicted Means

The PMN method is a two-step process. The first step is the modeling step, followed by a hot-deck step where imputed values replace missing multiplicities. The different attributes of the six multiplicity models, corresponding to the six pair domains, are described in this subsection.

Response categories. The response categories for the six multiplicity final response models were simply the multiplicity counts for each domain among the complete data cases.

Covariates in models. The pool of covariates for the response propensity models was the same pool that was used for the pair relationship response propensity models. By the same token, this pool also was used for the final response multiplicity models when the household composition age count variables were missing.[23] When these variables were not missing, the same pool again was used as with the pair relationship models. Naturally, the final set of covariates differed from the initial pool. The final set of covariates that were used in the models is provided in Appendix Q.

Building of models. For the child-focus parent-child domains, the count being modeled was the number of parents. In most cases, since the pair relationship had already been established, only two responses were possible within the parent-child pair relationship: one parent or two parents. There were rare instances where three parents could live in the household,

---

[22] In fact, multiple spouses were observed in the 2005 survey but were exceedingly rare.

[23] The widowed, divorced, and never married categories for marital status were combined into a single level for the multiplicity models.

with some combination of biological and step, foster, or adoptive parents. (Usually three parents were present when a stepparent lived in the house with the two biological parents.) For the purposes of modeling, the rare instances with more than two parents in the household were collapsed with the two-parent households. For these multiplicity counts, the fitted models were binomial logistic regression models. Only respondents who had a nonimputed pair relationship with nonmissing multiplicity counts was eligible for the model-building dataset.

The other responses (parent-focus parent-child and sibling-sibling multiplicity counts) were counts, where Poisson regression models were used. However, the data were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance.

Determination of predicted means. Although models were built using respondent pairs where the multiplicity was known definitively, predicted means were required for all pair domains where imputation was required. Once the models were fitted, predicted means were determined for both respondent pairs and nonrespondent pairs, using the parameter estimates from the models.

6.3.3.3 Constraints on Hot-Deck Neighborhoods and Assignment of Imputed Values

In the same manner as with the pair relationship imputations, donor pairs in the hot-deck step of PMN for these multiplicity domains were chosen with predicted means, if possible, within delta of the recipient pair's predicted mean. The value of delta varied depending on the value of the predicted mean. The values of delta for predicted probabilities are shown in Table 6.8.

Wherever necessary and feasible, logical and likeness constraints (as defined in Section 6.2.4.3) were placed on the membership in the hot-deck neighborhoods. The hot-deck step and the accompanying constraints are described separately for each of the variables in turn.

Parent-child pairs, child focus. The donor pairs and recipient pairs had to have the same pair relationship, excluding the restrictions on ages. This acted as a logical constraint. (Donor pairs had to have nonimputed pair relationship data.) In addition, the number of parents was restricted by the number in the household of the appropriate age. An additional constraint, therefore, was that donor pairs and recipient pairs had to have the same number of individuals in the household aged 26 or older, provided this information was available for the recipient pair. (Donor pairs had to have complete data on all the household composition age count variables.) If the recipient pair had only one person in the household in this age range, then the number of parents in the household could still have been two, if the other parent was younger than 26 years old. However, this constraint ensured that donor pairs and recipient pairs had the same household age pattern. This was a likeness constraint that was never loosened. Besides delta, additional likeness constraints involved the household composition.

In addition to the 26-or-older constraint, the neighborhoods were further restricted by requiring that donor pairs and recipient pairs have the same number of household members within the age ranges of 26 to 34, 35 to 49, and 50 or older. Other likeness constraints included requirements that donor pairs and recipient pairs have the same (1) number of household members younger than 12 years of age; (2) number of household members between the ages of 12 and 17 (inclusive); (3) household sizes; and (4) values for IRPRREL. This latter constraint strengthened the requirement of matching pair relationships to include the restrictions on the ages. It meant that, for example, donor pairs and recipient pairs within the domain involving 12- to 17-year-olds both involved 12- to 14-year-olds or both involved 15- to 17-year olds.

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor pair with the closest predicted mean or means; (2) abandon the requirement that donor pairs and recipient pairs had to have the same number of household members younger than 12 and between 12 and 17 (inclusive); and (3) abandon the requirement that donor pairs and recipient pairs had to have the same number of household members within the age ranges of 26 to 34, 35 to 49, and 50 or older, and drop the household size constraint. The IRPRREL constraint was never loosened.

Parent-child pairs, parent focus. As with the child-focus pairs, donor pairs and recipient pairs had to have the same pair relationship, and donor pairs were required to have nonimputed pair relationship data. For the parent-focus pairs, the counts could have taken on more than two values. If the counts from the two pair members did not get reconciled but both pair members had valid rosters, then the two counts acted as upper and lower bounds for the imputation, acting as additional logical constraints. The counts were limited anyway, however, since the age ranges of the children were, by definition, constrained. Specifically, donor pairs and recipient pairs had to have the same number of household members within the relevant age ranges (12 to 14, 12 to 17, 15 to 17, or 12 to 20, depending upon the recipient pair's value for IRPRREL). (As before, donor pairs had to have complete data on the roster age variables.) Additional likeness constraints included a requirement that donor pairs and recipient pairs have the same number of household members younger than 12 and a requirement that household sizes be the same. The constraint on IRPRREL also was included.

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor pair with the closest predicted mean or means; (2) abandon the requirement that donor pairs and recipient pairs had to have the same number of household members younger than 12 years old; (3) abandon the requirement that donor pairs and recipient pairs had to have the same household size; and (4) loosen the IRPRREL constraint.

Sibling-sibling pairs. As with the parent-child pairs, donor pairs and recipient pairs had to have the same value for IRPRREL, and donor pairs were required to have nonimputed pair relationship data. As with the parent-child parent-focus pairs, the counts from the two pair members acted as upper and lower bounds for the imputation, as additional logical constraints, provided both pair members had valid rosters. Donor pairs and recipient pairs also were required, as a logical constraint, to have the same number of household members within relevant age ranges. For example, for a sibling-sibling pair with ages 12 to 14 and 15 to 17, with a focus on the younger member, the donor pair and recipient pair were required to have the same number of 15- to 17-year-olds. (As before, donor pairs had to have complete data on the roster age variables.) Additional likeness constraints included a requirement that donor pairs and recipient pairs have the same (1) number of household members younger than 12; (2) household sizes; (3) number of household members in the age group corresponding to the pair member of focus; and

(4) number in the household between ages 12 and 17, for the sibling-sibling pairs where one member was between 12 and 14 (inclusive) and the other was between 15 and 17 (inclusive).

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor pair with the closest predicted mean or means; (2) abandon all likeness age count constraints; and (3) abandon the requirement that donor pairs and recipient pairs had to have the same household size.

6.3.3.4 Additions to the Analytic File

The imputation-revised versions of the parent-child multiplicity variables were called IRMPCCxx and IRMPCPxx, where the final C and P refer to the focus in the domain. The "xx" refers to the age range of the children, which is the upper bound if the lower bound is 12 or "57" if the range is 15 to 17. The edited version of these variables, MCPCCxx and MCPCPxx, also were released to the analytic file. The sibling-sibling imputation-revised variables were called IRMSxxyy, where the "yy" refers to the upper bound of the age range corresponding to the focus pair member, and the "xx" refers to the upper bound of the age range corresponding to the remaining pair member. The edited version of these variables was given by MCSxxyy. The imputation indicators also were released to the analytic file, with II prefixes instead of IR prefixes. Finally, the spouse-spouse counts were called MCSPSP and MCSPSPWC. These were simply indicators of whether the pair was a spouse-spouse pair, or whether the pair was a spouse-spouse pair with children younger than 18. No imputation was required for these variables.

6.4 Stage Three: Creation and Imputation of Household-Level Person Counts in Each Domain for the Purposes of Pair Weight Calibration

In order to improve the quality of the estimates from the pair data through poststratification, it was necessary to identify the household-level person counts for each domain. This entailed finding the number of individuals in the household that belonged to a particular domain, given one member of a domain was selected as the focus. These counts were more difficult to derive than the multiplicity counts since all households were considered. Within each household, counts for any of the domains of interest were derived, regardless of whether that household belonged to that domain, or even whether a pair was selected at all. The counts were derived for 10 of the 14 pair domains described in Section 6.3. For two of the remaining domains—the parent-child counts where the child was between 15 and 17—calculating the household counts was unnecessary.[24] For the other two remaining sibling-sibling domains, the reasons are historical: they were added after the procedures were first developed, and there was insufficient time to develop the household counts for those domains. The domains where these counts were created are listed below:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;

[24] Since household counts were defined for everybody, it was possible to derive these counts using the counts for the parent-child domains where the child was between 12 and 14 and where the child was between 12 and 17. The multiplicity counts for the parent-child (15 to 17) domain had to be calculated, however, and could not have been derived in this easy way. This was due to the fact that multiplicity counts were only defined if the pair relationship corresponded to the pair domain of interest.

3. parent-child (child 12 to 17), parent focus;
4. parent-child (child 12 to 17), child focus;
5. parent-child (child 12 to 20), parent focus;
6. parent-child (child 12 to 20), child focus;
7. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
8. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
9. spouse-spouse (includes partner-partner) with children younger than 18; and
10. spouse-spouse (includes partner-partner).

Determining the household-level person counts was a three-step process: (1) determine the household count for each respondent, whether a member of a pair or a single respondent; (2) use the screener, quality of roster, and other means to figure out the appropriate final count, either by attempting to reconcile differing counts between pair members or by attempting to determine the appropriate count when information from only one roster was available; and (3) impute missing counts. For households where only one respondent was selected, the matching step (step 2) was unnecessary. The first step is described in Section 6.4.1, the second in Section 6.4.2 and Appendix S, and the third in Section 6.4.3.

Since the pair weights reflected selection done at the time of screening, the household-level person counts should have reflected the household makeup at that time. As with the multiplicity counts, however, this was not entirely possible, so no account was made for cases where a change in the household makeup occurred between the screening time and the time of both interviews. An explanation for why this was not possible for the multiplicity counts is described in the introduction to Section 6.3. Moreover, as stated in that section, to implement such an adjustment would have been extremely complicated for the household-level person counts. Nevertheless, in cases where there were disagreements between pair members on the value of the household-level person count, the screener was used to resolve those disagreements.

6.4.1 Determining the Household-Level Person Count for Each Respondent

The multiplicity count was a count of the number of pairs in the household that could be associated with the person of focus. The household-level person counts asked a different question: How many persons of focus were there for a given pair domain, provided such a pair domain existed in the household, regardless of what pair (or whether a pair) was actually selected? For a parent-child pair, for example, if two parents were in the household with three children aged 12 to 14, then the household person count for the parent focus was the same as the multiplicity count for the child focus: 2. Similarly, the household person count for the child focus is the same as the multiplicity count for the parent focus. Household person counts also would have been obtained for the various sibling-sibling and spouse-spouse domains in this example, even though the relationship was parent-child.

6.4.1.1 Parent-Child Domains

When obtaining household-level person counts for parent-child domains, the six parent-child domains listed in the introduction to this section are what were under consideration. In any household, the household-level person counts for parent-child domains were nonzero if at least one parent was present in the household with children within the relevant age range. In this instance, the child-focus count would have been simply the number of children in the household within that age range that belonged to the parent in the household, and the parent-focus counts would have been the number of parents. If more than one "family unit" (mother and/or father with children) lived within the household, the child-focus counts should have counted children from more than one set of parents, and the parent-focus counts should have counted two or more parents, at least one for each set of children. One situation where this occurred was where three generations lived within the same household, with children in both the youngest and the second generations within the relevant age range. Using the youngest generation as the reference point, some of the parent's siblings (the grandparents' other children) were within the relevant age range. In this instance, the parent-child domains of the number of children would have included both children of the parents and the children of the grandparents who were in that age range. The count of the number of parents included both the parents and grandparents (and exceeded 2). Identifying more than one family unit in a household with children within the relevant age range under other scenarios (e.g., two sisters both with children within the relevant age range, both living within the same household) could not be determined from the data and had to be disregarded. Regardless of how many family units were in the household, counts had to be determined in different ways depending upon whether a parent-child pair "of interest" was selected or not. Descriptions of how to obtain the household-level person counts are provided below for the parent-child domains outlined above, first for parent-child pairs of interest, with parent-focus and child-focus domains considered together. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This was followed by descriptions for other pairs and single respondents, with parent-focus and child-focus domains considered separately.

6.4.1.1.1 Obtaining Counts for Parent-Child Domains (Parent-Focus and Child-Focus): Parent-Child Pairs, Child Younger Than 21

If the pair was identified as parent-child and the three-generation situation described above was not apparent, the household-level child-focus person count was given by the parent-focus multiplicity count. Similarly, the household-level parent-focus person count was given by the child-focus multiplicity count. If a three-generation situation was identified and the grandparent also had children within the relevant age range, the number of children and the number of parents were adjusted appropriately. The final household count in this instance was greater than the imputation-revised multiplicity count, which did not include all of the children in the household within the relevant age range.

6.4.1.1.2 Obtaining Counts for Child-Focus Parent-Child Domains: Other Pairs and Single Respondents[25]

For other pairs and single respondents, the following conditions were required to determine the household count of the number of children of parents in the household:

[25] "Other pairs" included pairs that were not within a domain of interest because the age of at least one of the pair members was outside the relevant age range. For parent-child pairs, this applied to a pair with a child that was 21 or older. For sibling-sibling pairs, this applied to siblings where both were within the same age range (both were 12 to 14, 15 to 17, or 18 to 25) or at least one of the siblings was older than 25 years of age. "Other pairs" also are referenced in Sections 6.4.1.1.3 and 6.4.1.2.2.

1. If the age of the respondent was within the relevant age range and that child had at least one parent, then the child-focus counts were determined in the same way as the parent-focus multiplicity counts: The count was of the self plus the child's siblings within the relevant age range. If the child's parents were not identified as living with him or her in the household, the count was set to 0.

2. If the respondent had children within the relevant age range, then the count was of the respondent's children within that range. If the respondent also had older children who had children of their own within the relevant age range, then the count was of the respondent's children and grandchildren within the relevant age range.

3. If the age of the respondent was outside the relevant age range but the respondent had parents living with them in the household and had siblings within the relevant age range, then the count was of the number of the respondent's siblings.

4. If the respondent had grandchildren within the relevant age range and the respondent also had children older than 25 or children-in-law living with them, then the count was the number of the respondent's grandchildren. (The assumption was that the respondent's children or children-in-law were the parents of the respondent's grandchildren. The likelihood of this not being the case was small. In the case where a pair was selected, this was resolved by looking at the count of the other pair member.)

6.4.1.1.3 Obtaining Counts for Parent-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count of the number of parents of children in the household:

1. If the age of the respondent was within the relevant age range, then the count was of the number of the respondent's parents (which could be 0).

2. If the age of the respondent was outside the relevant age range but the respondent had siblings within the relevant age range, then the count was of the number of the respondent's parents (again, this could be 0).

3. If the respondent had children within the relevant age range, then the parent-focus counts were determined in the same way as the parent-focus multiplicity counts: The count was of the self plus the spouse or live-in partner. If the respondent also had older children (older than 25 and living with him or her) who had children of their own (identified as grandchildren) within the relevant age range, then the count was at least 2. If the respondent had a spouse or live-in partner in the household, the count was incremented by 1, and if a child-in-law was in the household, the count also was incremented by 1. (Note that, under these scenarios, the number of parents could range between two and four.)

4. If the respondent had grandchildren within the relevant age range but no children in that range, and the respondent had a child older than 25 or a child-in-law living with

them, then the count was 2 if both the child older than 25 and the child-in-law were living in the household, 1 if not.

6.4.1.2 Sibling-Sibling Domains

When obtaining household-level person counts for parent-child domains, the two sibling-sibling domains listed in the introduction to this section are what were under consideration. As with the parent-child counts, the household-level person counts for sibling-sibling domains were nonzero if at least one sibling-sibling pair was present in the household within the relevant age ranges, in which the count was simply the number of appropriately aged siblings. If sets of siblings from more than one "family unit" (sets of siblings from different parents) resided within the same household, the sibling-sibling counts should have counted possible pairs from within each set. However, sets of siblings that did not involve the respondent's family unit could not have been identified from the data. Regardless of how many sets of siblings were in the household, counts had to be determined in different ways depending upon whether a sibling-sibling pair "of interest" was selected or not. Descriptions of how to obtain the household-level person counts are provided below for the sibling-sibling domains outlined above, first for sibling-sibling pairs of interest. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This was followed by descriptions for other pairs and single respondents. In each case, the descriptions apply regardless of which sibling-sibling domain was under consideration.

6.4.1.2.1 Obtaining Counts for Sibling-Sibling Domains: Sibling-Sibling Pairs of Interest

If the pair was identified as sibling-sibling within a relevant domain, the multiplicity count was simply given by the number of younger siblings since the older sibling was the focus. The household-level sibling-sibling person counts were determined in a similar manner to the multiplicity count, except that the count of interest was of the number of older siblings. If the pair member was the older sibling, then the household count was the self plus the number of siblings in the older age range. The count for the younger sibling pair member was simply the number of siblings within the same older age range. Unlike the case with the parent-child household-level counts, inconsistencies in the sibling-sibling counts when the pair selected was sibling-sibling still needed to be resolved. However, the rules for resolving inconsistencies followed directly from those used for the multiplicity counts when counting the number of younger siblings, provided in Appendix R. Note that a pair that was within one sibling-sibling pair domain had to be outside the other sibling-sibling pair domain.

6.4.1.2.2 Obtaining Counts for Sibling-Sibling Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count of the number of siblings within the older age ranges of the domains of interest in the household:

1. If the age of the respondent was within the age range of the older sibling and that child had at least one sibling in the younger age range, then the counts were given as

the self plus the child's siblings within the older age range. If the child did not have any siblings within the younger age range, the count was set to 0.

2. If the age of the respondent was within the age range of the younger sibling and that child had at least one sibling in the older age range, then the counts were given by the number of child's siblings in the older age range.

3. If the age of the respondent was outside the age range of the older or younger sibling but had at least one sibling in each of the older and younger age ranges, the counts were given by the number of siblings in the older age range.

4. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had children both within the older and the younger age ranges, then the count was of the number of respondent's children in the older age range.

5. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had grandchildren within the older and younger age ranges, then the count was of the number of grandchildren in the older age range. (If the respondent's grandchildren were cousins rather than siblings, there was no way of deciphering this from the data. This had to be resolved by looking at the information from the other pair member, if another pair member was selected.)

6.4.1.3 Spouse-Spouse Domains

What is referred to as a "spouse-spouse domain" was actually derived from spouse-spouse and partner-partner pair relationships. The following conditions were required for the number of spouse-spouse (including partner-partner) pairs to be incremented by one. Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair.
2. The respondent was not part of a spouse-spouse pair but had a spouse (or live-in partner).
3. The respondent had two parents living in the house.
4. The respondent had two parents-in-law living in the house.
5. The respondent had two grandparents living in the house.
6. The respondent had a child and a child-in-law living in the house.

The following conditions were required for the number of spouse-spouse pairs with children younger than 18 to be incremented by one. (These also include partner-partner pairs with children younger than 18.) Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair with children younger than 18.

2. The respondent was not part of a spouse-spouse pair[26] but had a spouse (or live-in partner) and children younger than 18.
3. The respondent had two parents living in the house and was either younger than 18 or had siblings younger than 18.
4. The respondent had a child and a child-in-law living in the house and had grandchildren younger than 18.

6.4.2 Determining the Final Household-Level Person Count

For a particular type of household-level person count, there are three types of households from a sample selection perspective. For the first type, a pair was selected and both pair members responded, where the pair relationship corresponded directly to the pair domain being counted. In this case, the household-level person count was usually easy to obtain using the multiplicity counts, although an adjustment was sometimes required if more than one family unit was in the household. For example, a parent-child pair was selected where the child was 12 years old, and the household-level person count for the parent-focus parent-child (12 to 14) domain was required. In the second type of household, a pair also was selected and both pair members responded, but in this type the pair relationship did not correspond directly to the pair domain being counted. In this case, determining the final count was sometimes more difficult, particularly if one or more of the counts was a count of 0. A count of 0 from a roster with good data did not necessarily mean that the final count should be 0. For example, suppose a household consisted of a man, his wife, brother, and two sons, and suppose one of the sons and his uncle (the man's brother) were selected. If the uncle's roster had a count of 0 for all domains of interest—since all of the household members were "other relatives" from his perspective—then no nonzero parent-child count could be obtained. The final count would have to be determined from imputation. In the third type of household, only one respondent was selected. In this case, it was not necessary to match counts from different pair members, but determining the final count could still be difficult if the count was 0 for a household where the value was not truly 0.

For situations where a pair was selected and both pair members had good roster data, if the counts agreed between the pair members and were not 0, then an easy determination of the final household-level count was possible. Surprisingly, this occurred in a majority of cases. If one pair member had a bad roster with no information in it and the other had a good roster, this was treated in the same way as if a single respondent was selected with a good roster. In either of these cases, the final count could be determined, provided a considerable number of conditions were satisfied. The conditions used to accept a good roster's count, when either the other pair member's roster was bad or no pair was selected, are provided in Appendix S. If these conditions were not met, the final household-level person count was left to imputation. Imputation also was required if two pair members were selected, both with bad rosters.

For the remainder of cases, some could be reconciled and some could not. In the cases where reconciliation was possible, some of the disagreements were caused by the pair members' rosters having different age and gender compositions. In these cases, many of the disagreements

[26]All spouse-spouse pairs were excluded here since spouse-spouse pairs with children were already accounted for, and spouse-spouse pairs without children had already been defined, possibly by imputation, not to have children younger than 18.

between the pair members were resolved by going to the screener. However, the screener did not provide much help if the age and gender composition of the pair members' rosters were identical, yet the counts still disagreed, as was the case with the nephew-uncle pair described above. In that example, one count was 0 and the other was nonzero. Under conditions set out in Appendix S, it was possible to determine that the disagreement in this case was due to the uncle not being able to identify the parent-child domains, and the nonzero count was used. More detailed rules for reconciling differences between pair members are described in Appendix S.

If the attempt to reconcile differences in the household-level person counts between pair members was unsuccessful, upper and lower bounds within which the imputed value must reside were determined from the counts for each pair member and the counts for the screener.

6.4.3 Creation of Imputation-Revised Household-Level Person Count Variables

Because of the difficulty in definitively determining household-level counts in many cases, imputation was not an uncommon proposition. As with the imputation of pair relationships and multiplicities, the imputation was conducted using the PMN method described in Appendix N. In this section, the application of PMN to the imputation of household-level person counts is described. Since only the household-level person count in the third stage was imputed for each household, the imputation was univariate in the sense that no sequential models were necessary. However, in some cases several variables were associated with a single model, as described below.

6.4.3.1 Setup for Model Building

Household-level person counts of the domains listed in the introduction to Section 6.4 were defined for all respondents, regardless of what pair they belonged to, or even whether they were within a pair at all. Moreover, since a nonzero count did not depend upon the respondent being within the relevant age range, no logical constraints on age were necessary. However, the age of the respondent did have an impact on the final count. The biggest difference in the presence or absence of particular domains in a household was the presence of youths younger than 18. This was especially true if there were two or more youths in a household, in which case the household-level person counts would be considerably different from situations where this was not the case. As a result, both the pair and single-respondent samples were split by age. For the pairs, both pair members in one sample were younger than 18, and the remainder of pairs were in the other sample. For the single respondents, one sample consisted of respondents younger than 18, and the other consisted of the remainder. Separate imputations were conducted in the two samples.

Four separate imputations were conducted for the sibling-sibling domains, arising from four separate models. Unlike the multiplicity counts, no imputations were conducted for the younger focus sibling-sibling domains. Hence, only two of the sibling-sibling domains had household-level person counts imputed. However, four separate imputations were required since the sample was split into two subsamples for both pairs and single respondents.

The parent-child domains were hierarchical, so as with the multiplicities, the imputations could not have been conducted independently if consistency was to be maintained. Hence, like the multiplicities, only two models were fitted to the child-parent pairs using just the counts for children (12- to 20-year-olds). One set of models was for the number of the children who had at least one parent, and the other set was for the number of parents who had a child aged 12 to 20. Using the predicted means from these models, a single donor pair was selected from which the household-level person counts were determined for 12-to-14, 12-to-17, and 12-to-20 child-parent pair domains. (The household-level person counts for the 15-to-17 child-parent domains were not determined but could be easily derived.) Since the household-level person counts for specific domains were not dependent upon the pair relationship, it was not necessary to impute the parent-focus and child-focus counts separately as was done with the multiplicities. Hence, although separate models were fit to the parent-focus and child-focus counts, the predicted values from these models were brought together in a single multivariate imputation.

The spouse-spouse household-level person counts also were hierarchical in that knowledge of whether a spouse-spouse pair was in the household was required before one could say that the pair had children. It was somewhat more complicated than the parent-child hierarchical setup, however, as one model could not represent whether there was a spouse-spouse pair in the household and whether that pair had children. As a result, the imputations were conducted in two stages, with the spouse-spouse pair imputations processed first, followed by the imputations of whether the pairs had children.

The first step for these models was to define respondents, nonrespondents, and the item response mechanism. For a pair or single respondent to be considered complete, the household-level person counts had to be nonmissing for all the variables being imputed. For the parent-child pair domains, this meant that the household-level person count had to be nonmissing for the parent-focus and child-focus 12-to-20 domains. Nonmissing household-level person counts for these domains automatically guaranteed nonmissing counts for the subset parent-child domains. A single response propensity adjustment was calculated for all the parent-child domains within each subsample, and separate response propensity adjustments were calculated for the remainder of domains. Separate response propensity adjustments were calculated for pairs and single respondents. For pairs, these adjustments were calculated in order to make the household weights representative of the entire sample of pairs. For single respondents, household weights also were used. The adjustments were calculated in order to make the respondent household weights representative of the entire sample of households that were not part of a pair. Because the spouse-spouse imputations were conducted in two stages, the response propensity adjustment for the spouse-spouse-with-children domain adjusted weights to be representative of all spouse-spouse pairs. Missing counts for the spouse-spouse-with-children domain were not imputed until it was known definitively, after the hot-deck step of the PMN imputation, whether a household had spouse-spouse pairs.

6.4.3.2 Model Building and Determination of Predicted Means

The PMN method is a two-step process. The first step is the modeling step, followed by a hot-deck step where imputed values replace the missing household-level person counts. The different attributes of the models are described in this subsection.

Response categories. The response categories for the household-level person count final response models were simply the household-level person counts, corresponding to each domain, among the complete data cases. In some cases, two family units were in a household. If these resulted in unusual household-level person counts, they were excluded from the modeling step, and were considered nonrespondents for the purposes of weight adjustment. No predicted mean was calculated in these cases. This occurred with the parent-child parent focus counts and the spouse-spouse-with-children counts. For the parent-child parent-focus counts, two family units sometimes resulted in counts of 3 or 4 parents, which were extremely rare levels. The response categories for the models in the case of the parent-child parent focus counts were, therefore, limited to 0, 1, or 2 or more. With the spouse-spouse-with-children counts, having two spouse-spouse pairs with children younger than 18 also was an extremely rare category. The response categories that resulted for the spouse-spouse-with-children models were, therefore, 0 or 1 or more. Households with two family units did not need to be excluded from the spouse-spouse models, since having two spouse-spouse pairs in a household, though not common, was not rare.

Covariates in models. The same pool of covariates that was used for the multiplicity models also was used for the household-level person counts. The same dual set of models were fitted according to whether the household composition age count variables existed or not. Naturally, the final set of covariates differed from the initial pool. The final set of covariates that were used in the models is provided in Appendix Q.

Building of models. The household-level person counts could have a value of 0, which distinguished them from the multiplicities from a modeling point of view. For the parent-focus parent-child domains, the count modeled was the number of parents, which had three values for reasons explained earlier: 0, 1, or 2 or more. The model for spouse-spouse pairs also had three levels: 0, 1, or 2 or more. Both of these models (within each subsample) were fitted as multinomial logistic models. Also for reasons stated earlier, the spouse-spouse-with-children models had only two levels (0 or 1 or more), so binomial logistic models were fitted to those data. Poisson regression was used to fit the models for the household-level person counts corresponding to the sibling-sibling domains, as well as the child-focus parent-child domains. The data were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance.

Determination of predicted means. Although models were built using respondent pairs and single respondents where the household-level person counts were known definitively, predicted means were required for all pairs and for all respondents who were not part of a pair. Once the models were fitted, predicted means were determined for respondent pairs and single respondents, as well as item nonrespondents among pairs and singles, using the parameter estimates from the models.

6.4.3.3 Constraints on Hot-Deck Neighborhoods and Assignment of Imputed Values

In the same manner as the multiplicity and the pair relationship variables, donors (among pairs and single respondents) in the hot-deck step of PMN for the counts associated with this domain were chosen with predicted means, if possible, within delta of the recipient's (whether a pair or single respondent) predicted mean. The value of delta varied depending on the value of the predicted means. The values of delta for predicted probabilities are shown in Table 6.8.

Wherever necessary and feasible, logical and likeness constraints (as defined in Section 6.2.4.3) were placed on the membership in the hot-deck neighborhoods. The hot-deck step and the accompanying constraints are described separately for each of the variables in turn.

In those instances where an imputed value could not be found after loosening all the likeness constraints, the imputed value was determined by doing a random imputation within bounds derived from the household composition. One of the situations where this occurred was when the household had two or more family units in the household. Even though the counts were not included in the models, no predicted means were calculated. (This occurred with the parent-focus parent-child counts, as well as the spouse-spouse-with-children counts.) Hence, instead of matching donors and recipients using predicted means, the imputed value was determined using the random imputation described earlier. Even though two-family households were included in the model for the child-focus parent-child counts, the resulting predicted means were not used. This was due to the fact that the parent-focus parent-child counts were in the same multivariate set as the child-focus parent-child counts, and the predicted means could not be used in the imputation of the parent-focus parent-child counts when two families were in the household.

6.4.3.3.1 Parent-Child Counts

Since parent-focus and child-focus counts were so closely related, a logical constraint was placed on donors such that if the parent-focus count was nonmissing and nonzero, then the child-focus count had to exceed 0. Similarly, a nonzero, nonmissing child-focus count required that the donor's parent-focus count exceed 0. If the child focus counts were missing, donors and recipients had to have the same number of household members in the age range corresponding to the domain of interest. (Donors had to have complete data on all the roster age variables.) The same constraint was applied if the parent-focus counts were missing but the child-focus counts were nonmissing, with an additional requirement: It had to be possible that no parent-child pairs existed in the household. (If it was known that there were parents in the household for the appropriate domain, it was not necessary to limit donors to have the same child age composition as the recipient.) These were likeness constraints that were never loosened. In addition, if a recipient had two family units in the household, a regular hot deck imputation could not be done, as stated earlier. For all missing counts, the counts from the two pair members (in the case of pair recipients) and the household composition were used to create upper and lower bounds, provided valid roster information was available. These bounds acted as additional logical constraints. Besides delta, additional likeness constraints all involved the household size and additional constraints on the household composition, which are described in the following paragraph.

An attempt was made to match donors and recipients in each of three age ranges that are commonly associated with children aged 12 to 20: 26 to 34, 35 to 49, and 50 or older. This likeness constraint was applied whether the child-focus or the parent-focus count was missing. However, its application in the case of a missing child-focus count and nonmissing parent-focus count required an additional condition: It had to be possible that no parent-child pairs existed in the household. (If it was known that there were children in the household who belonged to parents, it was not necessary to limit donors according to the parent age ranges.) A looser form of this constraint was to collapse the 26-to-34 and 35-to-49 age ranges into a single age range and drop the 50-or-older constraint. Other household composition constraints required donors and recipients to have the same number of household members younger than 12 years old and between the ages of 18 and 25 (inclusive).

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor with the closest predicted mean or means; (2) abandon the requirement that donors and recipients had to have the same household size; (3) abandon the requirement that donors and recipients had to have the same number of household members younger than 12, between 18 and 25 (inclusive), and 50-or-older, and collapse the 26-to-34 and 35-to-49 age constraints; and (4) remove the 26-to-49 age constraint.

6.4.3.3.2 Sibling-Sibling Counts

For all missing counts, the counts from the two pair members (in the case of pair recipients) and the household composition were used to create upper and lower bounds, provided valid roster information was available. These bounds acted as additional logical constraints. If the sibling-sibling counts were missing, donors and recipients had to have the same number of household members in the age range corresponding to the domain of interest. (Donors had to have complete data on all of the roster age variables.) Since imputations for the household-level person counts were done only on the sibling-sibling domains with the older sibling as the focus, this meant that donors and recipients had to have the same number in the household aged 15 to 17 (for the 12-to-14/15-to-17 domains) or aged 18 to 25 (for the 12-to-17/18-to-25 domains). This was a likeness constraint that could be loosened to a logical constraint: The imputed count could not exceed the recipient's number of household members in the relevant age range. An additional likeness constraint recognized the correlation between parent-child domains and sibling-sibling domains (i.e., the presence of parent-child domains in the household meant that a sibling-sibling domain would be more likely). Hence, donors and recipients both either had to have parent-child domains in the household or not have such domains. Other likeness constraints were related to the household composition.

In addition to matching donors and recipients on household size, they also had to match on the number of household members younger than 12, and the number of household members within the younger sibling's age range. The age constraints corresponding to the age ranges of the siblings could be loosened so that the counts for the donor and recipient for the older sibling's age range had to be both 0 or both nonzero.

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor with the closest predicted mean or means; (2) abandon the requirement that donors and recipients had to have the same household size and the same number of household members younger than 12; (3) abandon the requirement that donors and recipients had to have the same number of household members younger than 12 years old; and (4) remove age constraints corresponding to the age ranges of the siblings so that the only age constraint was logical: Donors' counts could not have exceeded the total number in the older sibling's age range. At the same time, the constraint that required donors and recipients have the same status with regard to parent-child domains was removed.

6.4.3.3.3 Spouse-Spouse Counts

For all missing counts, the counts from the two pair members (in the case of pair recipients) and the household composition were used to create upper and lower bounds, provided valid roster information was available. These bounds acted as logical constraints. In addition, if a recipient had two family units in the household, a regular hot-deck imputation could not be done, as stated earlier. The rest of the likeness constraints all used information from the household composition, with recognition of the fact that the vast majority of spouse-spouse pairs were male-female pairs.

For the spouse-spouse pairs, which also included partner-partner pairs, the constraints attempted to match donors and recipients as much as possible in their household age and gender pattern. This included some likeness constraints that were never loosened: Both donors and recipients were required to have the same number of household members younger than 18 if that number was 0, 1, or 2. If the recipient had two or more members in his or her household younger than 18, the donor also had to have two or more household members younger than 18. In addition, donors and recipients had to have the same number of household members aged 15 or older and the same number of males aged 15 or older.

Likeness constraints that were loosened also were related to the age and gender composition of the household. In particular, donors and recipients had to match their household size and the number of household members, as well as males, within the age ranges of 18 to 25, 26 to 34, 35 to 49, and 50 or older. Looser forms of these constraints required the same number of members in the household and the same number of males in the household within the age ranges of 18 to 34 and 26 to 49.

The likeness constraints were loosened in the following order (where applicable): (1) abandon the neighborhood and choose the donor with the closest predicted mean or means; (2) abandon the requirement that donors and recipients had to have the same household size and abandon the requirement that the donors and recipients had to have the same number of household members (male or female) within the age ranges of 18 to 25, 26 to 34, 35 to 49, and 50 or older; (3) abandon the requirement that donors and recipients had to have a similar number of household members younger than 18, as described above, and loosen the requirement that donors and recipients had to have the same number of males within the age ranges of 18 to 25, 26 to 34, and 35 to 49 so that donors and recipients were required to have the same number of males in the age ranges of 18 to 34 and 26 to 49; and (4) abandon the requirement that donors and recipients had to have the same number of males in the age range of 18 to 34.

6.4.3.3.4 Spouse-Spouse-with-Children Counts

The constraints for the spouse-spouse-with-children counts were exactly the same as the spouse-spouse constraints, with one exception. When the requirement that donors and recipients had to have a similar number of household members younger than 18 was abandoned, another constraint replaced it and was never loosened. This constraint required that if the recipient did not have anyone in the household younger than 18, then the same should be true of the donor. However, if the recipient did have someone in the household younger than 18, then the donor should too. Presumably this constraint was not necessary, since no imputation would be required if it was known that no children were in the household.

6.4.3.4 Additions to the Analytic File

The imputation-revised versions of the parent-child household-level person count variables were called IRHPCCxx and IRHPCPxx, where the final C and P refer to the focus in the domain. The "xx" refers to the age range of the children, which is the upper bound, since the lower bound is always 12. The edited version of these variables, HCPCCxx and HCPCPxx, also were released to the analytic file. The sibling-sibling imputation-revised variables were called IRHSxxyy, where the "yy" refers to the upper bound of the age range corresponding to the focus pair member, and the "xx" refers to the upper bound of the age range corresponding to the remaining pair member. The edited version of these variables was given by HCSxxyy. Finally, the imputation-revised versions of the spouse-spouse counts were given by IRHCSPSP and IRHCSPWC for the spouse-spouse and spouse-spouse-with-children counts, respectively. The edited versions of these spouse-spouse counts were called HCSPSP and HCSPSPWC. The imputation indicators also were released to the analytic file with II prefixes instead of IR prefixes.

7. Weight Calibration at Questionnaire Dwelling Unit and Pair Levels

The 2005 National Survey on Drug Use and Health (NSDUH) was based on probability sampling so that valid inferences can be made from survey findings about the target population. Probability sampling refers to sampling in which every unit on the frame is given a known, nonzero probability for inclusion in the survey. This is required for unbiased estimation of the population total. The assumption of nonzero inclusion probability for every pair of units in the frame also is required for unbiased variance estimation. The basic sampling plan involved four stages of selection across two phases of design: within Phase I, (1) the selection of census tracts within each State sampling (SS) region, (2) the selection of subareas or segments (comprised of U.S. Bureau of the Census blocks) within State sampling (SS) regions; (3) the selection of dwelling units (DUs) within these subareas; and, finally, within Phase II, (4) the selection of eligible individuals within DUs. Specific details of the sample design and selection procedures for the sample can be found in the 2005 NSDUH sample design report (Morton et al., 2006).

As part of the postsurvey data-processing activities, analysis weights that reflected the selection probabilities from various stages of the sample design were calculated for respondents. These sample weights were adjusted at the DU (screening sample), questionnaire dwelling unit (QDU), person, and paired respondent levels (the latter three all based around the drug questionnaire sample) to account for bias due to extreme values (ev), nonresponse (nr), and coverage.

The final sample weights for Phase I screener dwelling units (SDU) and Phase II QDU, person, and pair levels for the 2005 samples consisted of products of several factors, each representing either a probability of selection at some particular stage or some form of ev, nr, or ps calibration adjustment. In the following sections, we describe the QDU and pair weight components in greater detail. In summary, the first 10 factors were defined for all SDUs and reflected the fully adjusted SDU sample weight. The remaining components branched to reflect QDU and pair selection probabilities, as well as additional adjustments for ev, nr, and ps. Note that the final QDU and pair weights for the 2005 survey sample are the product of all weight components for each type of sample, illustrated in Exhibits 7.1 and 7.2.

For QDU data, generalized exponential modeling (GEM) calibration modeling was applied by partitioning the data into four groups of States: Northeast, South, Midwest, and West, based on census regions in the interest of computational feasibility. Previous experience showed that with current computing power, the large number of variables and records prevented any further reduction of modeling groups.

For pair data, GEM modeling was initially applied by partitioning the pair data into four groups based on census regions. However, there were not enough observations in each group to fit a comprehensive model to reduce bias. Alternatively, a single model was attempted for the whole pair data, but it was rejected as not practical due to computational limitations. A compromise approach was adopted by combining census regions into two groups: Northeast with South and Midwest with West. This grouping proved both manageable and desirable as it assisted in bias reduction, ease of modeling, and workload reduction. Exhibit 7.3 provides more details of the data partition for GEM modeling. The resulting sample sizes of selected and respondent units for the pair and QDU data partitions are shown for the 2001–2005 surveys in Table 7.1.

It may be noted that for the pair data in the 1999, 2000, and 2001 surveys, the built-in ev control feature of GEM was not used until the final respondent pair ev adjustment step. The reason for this is that the definition for ev domain was not finalized before the pair data calibration process was begun. However, for the 2002, 2003, 2004, and 2005 survey pair data, the built-in ev control feature was used for each adjustment step.

Exhibit 7.1 Summary of 2005 NSDUH QDU Sample Weight Components

Phase I Screener Dwelling Unit Level

Design Weight Components	
#1	Inverse Probability of Selecting Census Tract
#2	Inverse Probability of Selecting Segment
#3	Quarter Segment Weight Adjustment
#4	Subsegmentation Inflation Adjustment
#5	Inverse Probability of Selecting SDU
#6	Subsampling of Added SDU Adjustment
#7	SDU Release Adjustment

Weight Adjustment*	
#8	SDU Nonresponse Adjustment ***(res.sdu.nr)***
#9	SDU Poststratification Adjustment ***(res.sdu.ps)***
#10	SDU Extreme Value Adjustment ***(res.sdu.ev)***

Phase II Questionnaire Dwelling Unit Level

Design Weight Component	
#11	Inverse of Selection Probability of at Least One Person in the Dwelling Unit

Weight Adjustment*	
#12	Selected QDU Poststratification to SDU-based Control Totals ***(sel.qdu.ps)***
#13	Respondent QDU Nonresponse Adjustment ***(res.qdu.nr)***
#14	Respondent QDU Poststratification to SDU-based Control Totals ***(res.qdu.ps)***
#15	Respondent QDU Extreme Value Adjustment ***(res.qdu.ev)***

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

* These adjustments use the generalized exponential model (GEM), which also involves pre- and post-processing in addition to running the GEM macro. See Exhibit 4.1 (Chen et al., 2007). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 7.3.

Exhibit 7.2 Summary of 2005 NSDUH Person Pair Sample Weight Components

Phase I Screener Dwelling Unit Level

Design Weight Components	
#1	Inverse Probability of Selecting Census Tract
#2	Inverse Probability of Selecting Segment
#3	Quarter Segment Weight Adjustment
#4	Subsegmentation Inflation Adjustment
#5	Inverse Probability of Selecting SDU
#6	Subsampling of Added SDU Adjustment
#7	SDU Release Adjustment

Weight Adjustment*	
#8	SDU Nonresponse Adjustment ***(res.sdu.nr)***
#9	SDU Poststratification Adjustment ***(res.sdu.ps)***
#10	SDU Extreme Value Adjustment ***(res.sdu.ev)***

Phase II Person Pair Level

Design Weight Component	
#11	Inverse of Selection Probability of a Person Pair in SDU

Weight Adjustment*	
#12	Selected Pair Poststratification to SDU-based Control Totals ***(sel.pr.ps)***
#13	Respondent Pair Nonresponse Adjustment ***(res.pr.nr)***
#14	Respondent Pair Poststratification Adjustment to SDU-based Control Totals ***(res.per.ps)***
#15	Respondent Pair Extreme Value Adjustment ***(res.per.ev)***

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

* These adjustments use the generalized exponential model (GEM), which also involves pre- and post-processing in addition to running the GEM macro. See Exhibit 4.1 (Chen et al., 2007). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 7.3.

Exhibit 7.3 U.S. Bureau of the Census Regions/Model Groups

Model Group	Census Region
QDU	
1	**Northeast (9 States)**
	Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
2	**Midwest (12 States)**
	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
3	**South (16 States and the District of Columbia)**
	Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia
4	**West (13 States)**
	Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming
Pair	
1	**Northeast + South (25 States and the District of Columbia)**
	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
2	**Midwest + West (25 States)**
	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 7.1 Sample Size, by Model Group at QDU and Pair Levels

Model Group	2001 Selected QDUs	2001 Completed QDUs	2002 Selected QDUs	2002 Completed QDUs	2003 Selected QDUs	2003 Completed QDUs	2004 Selected QDUs	2004 Completed QDUs	2005 Selected QDUs	2005 Completed QDUs
QDU										
Northeast	14,208	11,155	11,436	9,724	11,639	9,732	11,466	9,552	11,599	9,617
South	19,814	16,029	17,121	14,877	17,194	14,676	17,200	14,712	17,579	14,744
Midwest	18,903	14,804	15,582	13,489	15,542	13,288	15,735	13,304	15,996	13,342
West	13,772	11,146	11,547	9,998	11,809	10,057	11,850	10,083	12,069	10,190
Total	66,697	53,134	55,686	48,088	56,184	47,753	56,251	47,656	57,243	47,893
Model Group	**2001 Selected Pairs**	**2001 Completed Pairs**	**2002 Selected Pairs**	**2002 Completed Pairs**	**2003 Selected Pairs**	**2003 Completed Pairs**	**2004 Selected Pairs**	**2004 Completed Pairs**	**2005 Selected Pairs**	**2005 Completed Pairs**
Pair										
Northeast + South	11,436	7,869	12,463	10,005	12,628	9,859	12,828	10,066	13,227	10,168
Midwest + West	11,612	7,926	12,432	10,033	12,819	10,172	12,894	10,043	13,355	10,247
Total	23,048	15,795	24,895	20,038	25,447	20,031	25,722	20,109	26,582	20,415

QDU = questionnaire dwelling unit

7.1 Phase I SDU-Level Weight Components

A total of 10 weight components for the SDU level correspond to selection probabilities and nonresponse, poststratification, and extreme value adjustment factors. Note that this differs from previous NHSDAs and NSDUHs in that a new design-based component was incorporated at the beginning of the process so that corresponding weight component numbers are incremented by one when compared to previous survey years with an otherwise similar weighting scheme. The first seven components in the Phase I sample weights reflect the probability of selecting the DUs. These components were derived from (1) the probability of selecting the census tract within each State sampling (SS) region, (2) the probability of selecting the geographic segment within each SS region, (3) a quarter segment weight adjustment, (4) a subsegmentation inflation factor, (5) the probability of selecting a DU from within each counted and listed sampled segment, (6) the probability of inclusion of added DUs, and (7) DU percent release adjustment. The three remaining weight components, #8 through #10, are GEM calibration adjustments accounting for (8) DU nonresponse at the screening level, (9) DU poststratification to census controls, and (10) DU-level extreme value adjustment, although in 2005 extreme value adjustment at this stage was deemed unnecessary, and thus Weight Component #10 was set to one for all respondent DUs. The person-level, QDU-level, and person pair-level weights use the product of the above 10 weight components as the common initial weight before further adjustments. For more detailed information on Weight Components #1, #2, and #4 through #7, refer to the 2005 NSDUH sample design report (Morton et al., 2006), and for more detail on Weight Components #3 and #8 through #10, see the 2005 person-level sampling weight calibration report (Chen et al., 2007).

7.2 QDU Weight Components

7.2.1 QDU Weight Component #11: Inverse of Selection Probability of at Least One Person in the Dwelling Unit

The selection of a questionnaire dwelling unit from all completed SDUs is based on the outcome of a variant of Brewer's method, which may select zero, one, or two persons. Any pair of survey eligible residents within the dwelling unit had some known, nonzero chance of being selected for the survey. The value for Weight Component #11 is equal to the inverse of the probability that at least one person in the dwelling unit is selected (see Section 2.2 for details).

7.2.2 QDU Weight Component #12: Selected QDU Poststratification to SDU-Based Control Totals

This poststratification factor adjusts the weights for selected QDUs to the SDU-based control totals. The SDU-based control totals are obtained by using the calibrated SDU weights. This adjustment step provides more stable controls for the subsequent nonresponse adjustment (Weight Component #13). Exhibit 4.1 lists the initially proposed variables for GEM modeling. The predictor variables are either 0/1 indicators or counting variables representing the number of persons who fall into a given demographic domain. The counting variables are derived from the screener demographic information. It may be noted that during screening, the only required demographic information was the age of each person rostered. Thus, other demographic information necessary for weight calibration, such as race/ethnicity and gender may be missing for certain rostered eligible persons, and so imputation was done to replace this missing data. For more details on the imputation of screener demographic information, see Chen et al. (2007).

The details on the predictor variables retained in the model and model summary statistics can be found in Appendix C.

7.2.3 QDU Weight Component #13: Respondent QDU Nonresponse Adjustment

This nonresponse adjustment step accounts for the failure to obtain respondent person(s) from each and every selected QDU. The same set of initially proposed predictor variables were used as for the previous adjustment (#12).

See Appendix C for more details on the predictor variables retained in the model and model summary statistics.

7.2.4 QDU Weight Component #14: Respondent QDU Poststratification to SDU-Based Control Totals

This final poststratification for all respondent QDUs utilized the same set of initially proposed predictor variables as previous adjustments. The corresponding control totals were obtained from the SDU-level sample, as was done for Weight Component #12.

See Appendix C for more details on the predictor variables retained in the model and model summary statistics.

7.2.5 QDU Weight Component #15: Respondent QDU Extreme Value Adjustment

The extreme weight proportions for the final poststratified weights were acceptably low, and so it was decided that the extreme value adjustment was not needed. Weight Component #15 was set to one for each responding QDU.

7.3 Pair-Level Weight Components

Exhibit 4.2 lists the initially proposed predictor variables for the following adjustment steps via GEM.

7.3.1 Pair Weight Component #11: Inverse of Selection Probability of a Person Pair in the DU

Selection of pairs of individuals from all eligible persons residing within the dwelling unit is based on the outcome of a variant of Brewer's method, which may select zero, one, or two persons. Any pair of survey eligible residents within the dwelling unit has some known, nonzero chance of being selected for the survey. When two persons are selected, a pair is formed. The pair selection probability is determined by the formula in Chapter 2. This weight component is the inverse of the selection probability discussed above.

7.3.2 Pair Weight Component #12: Selected Pair Poststratification to SDU-Based Control Totals

Similar to QDU Weight Component #12, this step was motivated by the consideration that the larger sample of all possible pairs provides more stable control totals for the respondent pair nonresponse adjustment. The weights of selected pairs were poststratified to the control totals that derived from calibrated SDU weights of all possible pairs. The pair-level demographic variables for all selected pairs, such as pair age group, pair race/ethnicity, etc., were derived from screener demographic information.

The details on the predictor variables retained in the model and model summary statistics can be found in Appendix H.

7.3.3 Pair Weight Component #13: Respondent Pair Nonresponse Adjustment

If both persons in the selected pair completed interviews successfully, the pair then was considered a respondent pair. This adjustment step accounts for failure to obtain respondent pairs from all selected pairs. In this step, respondent pair weights were adjusted to the control totals based on the full sample of selected pairs. Due to the low response rate of person pairs, this step had a relatively large adjustment on the weights. The same set of proposed predictor variables was used as for Weight Component #12. Similar to Weight Component #12, the pair level demographic variables for all selected pairs, such as pair age group, pair race/ethnicity etc., were derived from screener demographic information.

See Appendix H for more details on the predictor variables retained in the model and model summary statistics.

7.3.4 Pair Weight Component #14: Respondent Pair Poststratification to SDU-Based Control Totals

This final poststratification utilized the same set of initially proposed predictor variables as previous adjustment steps. In addition, 10 pair relationship domain-level indicator variables were added to the set of covariates. The control totals for GEM calibration were derived from the SDU sample of all possible pairs of eligible persons, as was done for Weight Component #12. The calibration control totals for these 10 domains used household-level person counts and the final QDU weights. As mentioned in the introduction, use of these household-level count totals for pair relationship domains in GEM calibration provided Hajek-type weight adjustment in the interest of obtaining more stable estimates. In setting up calibration covariates, multiplicity factors were needed. These factors, as discussed in the introduction, are used in constructing estimates for person-level parameters based on pair-related drug behavior. The factors depend on the pair domains of interest. For a selected set of pair domains, multiplicity factors are provided along with the pair-level analysis weights. See Chapter 6 for more detail on creation of and imputation of missing values in the pair relationship, multiplicity, and household-level person counts. See Chapter 4 for more detail on the use of multiplicities and household-level person counts in poststratification.

Unlike Weight Components #12 and #13, demographic covariates were based on data from the questionnaire instead of information pulled from the dwelling unit screener.

For more details on the predictor variables retained in the GEM model and model summary statistics, see Appendix H.

7.3.5 Pair Weight Component #15: Respondent Pair Extreme Weight Adjustment

We checked the extreme weight proportions for the weights up to Weight Component #14, using the extreme weight domains (see Section 5.2). Even though the previous adjustment steps utilized the built-in extreme weight control feature of GEM, the extreme weight proportions were still high enough to cause concern that they might produce unreliable estimates. Therefore, the extreme weight adjustment via GEM was implemented, using the same final set of predictor variables kept in the model for Weight Component #15. This step was successful in reducing the extreme weight proportion in all model groups. For details, see Appendix J.

8. Evaluation of Calibration Weights

During the weight calibration process, several criteria for quality control were implemented to assess model adequacy. In this chapter, we describe the individual procedures and a summary of their results. All tables referred to in this chapter can be found in Appendices D through G and I through L.

8.1 Response Rates

Table D.1 in Appendix D displays the final selected and responding questionnaire dwelling unit (QDU) sample sizes from the 2005 National Survey of Drug Use and Health (NSDUH) for various national domains. This table also shows the weighted response rates. Most domains reflect the overall 78.16 percent response rate, with most rates relatively close to 80 percent, although the highest response rate is 96.22 percent, from the household type 12-to-17 category. The lowest response rate came from the household type 26+ category, with 73.93 percent.

Table I.1 in Appendix I displays the final selected and responding pair-level sample sizes from the 2005 survey, for various national domains. Due to the nature of the pair data, the response rates were lower in all domains examined than at the QDU level, with an overall response rate of 70.37 percent. The response rates range from a low of 54.75 percent in the pair race/ethnicity Hispanic or Latino black or African-American category to a high of 90.58 percent from the group quarters domain. This extreme range of response rates is probably due to a combination of small sample sizes and response burden as a result of selection of pairs within households among various domains. Like at the QDU level, the top response rates are among the younger respondents (as measured by household type for the QDU data and pair age for the pair data). This pattern may be related to the relatively high response rates in the group level of the variable group quarters since it includes college dormitories.

8.2 Proportions of Extreme Value and Outwinsor Weights

During the stages of modeling adjustments (i.e., nonresponse [nr] and poststratification [ps]), one major issue of concern when deciding the adequacy of a particular model was the extent of the resulting proportion of extreme value (ev) and outwinsor weights (see Sections 5.1 and 5.2 for these definitions). For each weight adjustment step, these proportions are computed before and after the step for various domains. Prior to adjustment, the product of all weight components is used to compute proportions of evs and outwinsors, while after the adjustment the product includes the new adjustment factor. If the proportion of evs and outwinsors are deemed high, a separate ev treatment step after ps could be performed. This was done for the pair-level weights. Details of this step are explained in Section 7.3.5. A separate ev treatment step was deemed unnecessary for the QDU-level weights.

Tables E.1 and E.2 and Tables J.1 through J.3 present percentages of evs at the QDU level and the pair level, respectively, for various domains. Unweighted percentages are the percentage of actual counts of units defined as evs relative to the total sample size. Weighted percentages reflect the percentage of total ev weights relative to the total sample weight, while outwinsor percentages represent the total amount of residual weight when the weights are trimmed to the critical values (used for ev definition) relative to the total sample weight. For evaluation purposes, the outwinsor percentage is considered the most important of the three percentages, as this gave a measure of the impact of winsorization (or trimming) of ev weights (if we performed this treatment). See Sections 5.1 and 5.2 for the domains that were used to define extreme values.

8.3 Slippage Rates

The slippage rate for a given domain is defined as the relative percentage difference between the sampling weights and the external control totals, both before and after ps. The control totals for QDU and person pair ps are derived from the screener dwelling unit (SDU) weights, which were poststratified to U.S. Bureau of the Census population estimates (Chen et al., 2007). Table F.1 displays QDU national domain-specific weight sums for both before and after ps, as well as the desired totals to be met through ps. Table K.1 shows the same for the pair sample. These tables also show the relative percentage difference, or the amount of adjustment necessary (positive or negative) to meet the desired totals. The first relative difference is used explicitly during the ps modeling procedure to identify potential problems for convergence. Large differences in domains with relatively small sample sizes are indicative of potential large adjustment factors, which may cause problems in convergence while satisfying bound constraints. The reason is that adjustments required for one domain may have an adverse effect on another domain when a unit belongs to both.

As an example, consider that Table F.1, for the 2005 QDU domain household size of one, indicates a sample size of 5,673 with a total design-based weight of 29,590,028 and a census total of 29,857,979 with an initial slippage rate of -0.90 percent, which would imply a common weight adjustment of ≈ 1.009055, if this were the only calibration control. Similarly, looking at pair data in Table K.1, the pair domain category of pair age 18-25, 35-49 has a sample size of 1,119, a design-based weight of 17,066,849, and a census total of 16,965,106, showing an initial slippage of 0.60 percent. The resultant required adjustment would be ≈ 0.994039, if this were the only control. However, in the generalized exponential model (GEM), all controls are simultaneously satisfied under a complex algorithm that allows for different adjustment factors for different units.

8.4 Weight Adjustment Summary Statistics

Tables G.1, G.2, and L.1 through L.3 display summary statistics on the product of weight components before and after all stages of adjustment for the QDU and person pair, respectively. The summary statistics include sample size (n), minimum (min), maximum (max), median (med), 25th percentile (Q1), 75th percentile (Q3), and the unequal weighting effect (UWE). Note that in Tables L.2 and L.3 the sample size for pair age group, pair race/ethnicity, and pair gender are slightly different. This is because those variables were defined using screening demographic information in the nonresponse adjustment of respondent pairs, while in the poststratification of respondent pairs, they were defined from questionnaire demographic information. Because UWE is directly affected by weight adjustment factors and extreme weights, these values—along with the percentage of extreme weights as noted in Section 8.2—were used as guidelines for determining model adequacy.

8.5 Sensitivity Analysis of Drug Use Estimates

It is known that, in general, there is a trade-off between bias reduction and variance reduction. For instance, with GEM (for nr or ps), enlarging a simple model (such as the one with only main effects) has the potential of further reducing the bias. At the same time, this enlargement also may be associated with a corresponding increase in the variance of the estimate due to additional variability caused by estimating the model parameters. To check for possible overfitting of the GEM model, we conducted a sensitivity analysis for respondent QDU poststratification for the QDU weights, respondent pair poststratification, and extreme weight adjustment for the person pair weights. A simple baseline model was fitted with the same bounds and maximum number of iterations as was used for the chosen (more complex) final model. We then looked for substantial changes in point estimates and standard errors (SEs). For the QDU weights, some household-level characteristics were selected such as family income, number of youths, whether the household had health insurance coverage, and number of elders living in the household. The estimates and SEs are displayed in Table 8.1. For the person pair weights, selected licit and illicit drug use prevalence rates of 12- to 17-year-olds were calculated from parent-child pairs, and estimates and SEs of the estimates based on pair weights are shown in Tables 8.2a to 8.7b.

As seen in Table 8.1, the estimates and their SEs for the two models (baseline and the final) are generally similar to each other for the QDU weights. However, among the person pair estimates and SEs, there are some differences, but they do not seem significant in general.

Since the sensitivity analyses for both QDU- and pair-level calibrated weights seem to indicate that adding more covariates does not introduce an undesirable degree of instability in the estimates or their SEs, the final, more complex GEM models were deemed reasonable.

Table 8.1 Estimates of Totals and SEs for Domains of Interest Based on QDU Sample: 2005

	2005		
Domain	***n***	**Baseline**[1]	**Final**[2]
Households with Family Income			
$0-10,000	4,933	10,268,206 (327,560)	10,273,942 (328,087)
$10,000-20,000	6,312	15,538,013 (381,651)	15,537,085 (382,232)
$20,000-30,000	6,001	14,181,766 (351,957)	14,188,759 (353,641)
$30,000-40,000	5,707	13,444,387 (312,635)	13,441,625 (312,595)
$40,000-50,000	5,463	12,910,158 (307,397)	12,915,065 (307,333)
$50,000-75,000	8,235	18,897,884 (378,454)	18,895,079 (378,825)
$75,000-100,000	4,939	11,507,203 (301,237)	11,505,488 (301,424)
100,000+	6,303	15,888,733 (460,194)	15,879,309 (459,525)
Households with Number of Youths (< 18)			
0	19,511	71,234,543 (917,068)	71,237,230 (917,800)
1	11,582	17,280,868 (285,599)	17,278,181 (285,611)
2	10,043	15,274,993 (282,196)	15,277,059 (282,422)
3	4,493	6,232,030 (151,755)	6,227,950 (152,297)
4+	2,264	2,613,916 (83,687)	2,615,930 (84,378)
Households with Insurance Coverage			
Yes	38,515	95,270,605 (1,009,146)	95,260,756 (1,009,458)
No	9,378	17,365,745 (294,312)	17,375,594 (295,036)
Households with Number of Older Adults (65+)			
0	43,311	87,026,081 (818,400)	87,024,185 (818,308)
1	3,218	17,200,088 (526,204)	17,204,684 (526,182)
2	1,339	8,328,611 (309,184)	8,325,731 (309,250)
3+	25	81,570 (21,313)	81,751 (21,342)

QDU = questionnaire dwelling unit; SE = standard error.
Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last step of calibration, res.qdu.ps, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.2a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, by Mother Use: 2005

Drug	Mother User	2005 *n*	2005 Baseline[1]	2005 Final[2]
Alcohol				
Lifetime	Yes	2,196	40.64 (1.67)	40.71 (1.67)
	No	262	20.07 (3.79)	19.46 (3.60)
	Overall	2,458	37.76 (1.55)	37.78 (1.54)
Past Year	Yes	1,770	35.82 (1.86)	35.96 (1.86)
	No	688	18.63 (2.14)	19.21 (2.20)
	Overall	2,458	30.59 (1.46)	30.91 (1.46)
Past Month	Yes	1,306	19.46 (1.84)	19.85 (1.85)
	No	1,152	11.04 (1.36)	11.13 (1.39)
	Overall	2,458	15.34 (1.10)	15.61 (1.12)
Cigarettes				
Lifetime	Yes	1,750	26.89 (1.73)	27.46 (1.75)
	No	708	12.12 (1.84)	12.08 (1.86)
	Overall	2,458	22.16 (1.38)	22.63 (1.40)
Past Year	Yes	757	26.63 (2.58)	27.24 (2.59)
	No	1,701	10.73 (1.32)	11.07 (1.36)
	Overall	2,458	15.15 (1.24)	15.56 (1.26)
Past Month	Yes	684	18.79 (2.50)	18.91 (2.48)
	No	1,774	6.79 (1.04)	6.70 (1.00)
	Overall	2,458	9.82 (1.02)	9.78 (1.00)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.2b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, by Father Use: 2005

		2005		
Drug	**Father User**	***n***	**Baseline**[1]	**Final**[2]
Alcohol				
Lifetime	Yes	1,549	37.48 (2.04)	37.54 (2.04)
	No	104	22.53 (5.42)	22.33 (5.65)
	Overall	1,653	36.52 (1.96)	36.55 (1.97)
Past Year	Yes	1,291	30.25 (2.06)	29.78 (2.02)
	No	362	17.80 (2.94)	17.95 (3.02)
	Overall	1,653	27.49 (1.73)	27.18 (1.71)
Past Month	Yes	1,069	14.95 (1.88)	14.70 (1.81)
	No	584	8.57 (1.67)	8.71 (1.69)
	Overall	1,653	12.66 (1.33)	12.55 (1.29)
Cigarettes				
Lifetime	Yes	1,279	22.18 (1.74)	22.24 (1.74)
	No	374	12.77 (2.72)	11.92 (2.42)
	Overall	1,653	19.81 (1.54)	19.61 (1.51)
Past Year	Yes	533	21.71 (2.72)	21.57 (2.79)
	No	1,120	10.96 (1.41)	10.76 (1.34)
	Overall	1,653	13.88 (1.37)	13.71 (1.33)
Past Month	Yes	475	11.52 (2.13)	11.51 (2.18)
	No	1,178	5.25 (0.83)	5.21 (0.84)
	Overall	1,653	6.76 (0.80)	6.74 (0.81)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.3a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Mother-Child (12 to 17) Pairs, by Mother Use: 2005

Drug		Mother User	2005 *n*	2005 Baseline[1]	2005 Final[2]
Any Illicit					
	Lifetime	Yes	1,378	35.45 (2.10)	35.57 (2.09)
		No	1,080	16.53 (1.81)	16.37 (1.79)
		Overall	2,458	26.77 (1.47)	26.86 (1.45)
	Past Year	Yes	257	39.63 (4.65)	40.02 (4.74)
		No	2,201	18.39 (1.37)	18.34 (1.34)
		Overall	2,458	20.56 (1.35)	20.58 (1.34)
	Past Month	Yes	130	26.08 (6.35)	26.61 (6.57)
		No	2,328	10.63 (1.04)	10.53 (1.01)
		Overall	2,458	11.35 (1.05)	11.28 (1.03)
Marijuana					
	Lifetime	Yes	1,258	26.46 (2.11)	26.62 (2.11)
		No	1,200	9.16 (1.42)	8.88 (1.37)
		Overall	2,458	17.60 (1.33)	17.63 (1.32)
	Past Year	Yes	156	28.52 (6.39)	29.04 (6.52)
		No	2,302	14.02 (1.31)	14.01 (1.30)
		Overall	2,458	14.84 (1.26)	14.86 (1.26)
	Past Month	Yes	85	17.15 (7.00)	17.72 (6.93)
		No	2,373	7.27 (0.92)	7.25 (0.91)
		Overall	2,458	7.53 (0.92)	7.51 (0.90)

Note: Standard errors of prevalence estimates are provided in parentheses.

[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.

[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.3b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Father-Child (12 to 17) Pairs, by Father Use: 2005

		2005		
Drug	**Father User**	***n***	**Baseline[1]**	**Final[2]**
Any Illicit				
Lifetime	Yes	1,065	25.47 (2.18)	25.38 (2.21)
	No	588	16.77 (2.46)	17.00 (2.56)
	Overall	1,653	22.16 (1.58)	22.18 (1.63)
Past Year	Yes	197	24.92 (4.30)	24.88 (4.27)
	No	1,456	15.32 (1.49)	15.31 (1.54)
	Overall	1,653	16.32 (1.41)	16.30 (1.44)
Past Month	Yes	118	15.13 (5.31)	15.40 (5.40)
	No	1,535	7.41 (1.11)	7.52 (1.14)
	Overall	1,653	7.85 (1.09)	7.97 (1.13)
Marijuana				
Lifetime	Yes	1,001	16.44 (1.78)	16.11 (1.77)
	No	652	6.53 (1.30)	6.76 (1.48)
	Overall	1,653	12.18 (1.14)	12.07 (1.17)
Past Year	Yes	145	22.70 (5.10)	22.50 (5.08)
	No	1,508	8.94 (1.08)	8.86 (1.12)
	Overall	1,653	10.00 (1.08)	9.91 (1.10)
Past Month	Yes	96	14.98 (5.92)	15.17 (6.01)
	No	1,557	3.95 (0.69)	3.92 (0.70)
	Overall	1,653	4.49 (0.72)	4.48 (0.73)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.4 Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

Drug		Parent Talked about Dangers with Child	2005		
			n	**Baseline[1]**	**Final[2]**
Alcohol					
	Lifetime	Yes	2,522	34.42 (1.60)	34.60 (1.61)
		No	1,546	41.60 (2.13)	41.46 (2.16)
		Overall	4,068	37.26 (1.31)	37.30 (1.32)
	Past Year	Yes	2,522	27.13 (1.43)	27.48 (1.45)
		No	1,546	32.78 (1.97)	32.36 (1.96)
		Overall	4,068	29.36 (1.18)	29.41 (1.18)
	Past Month	Yes	2,522	13.04 (1.01)	13.17 (1.02)
		No	1,546	16.21 (1.47)	16.19 (1.44)
		Overall	4,068	14.29 (0.86)	14.36 (0.86)
Cigarettes					
	Lifetime	Yes	2,522	19.69 (1.27)	20.16 (1.31)
		No	1,546	23.71 (1.82)	23.59 (1.81)
		Overall	4,068	21.28 (1.07)	21.51 (1.08)
	Past Year	Yes	2,522	13.75 (1.17)	14.20 (1.22)
		No	1,546	16.35 (1.61)	16.23 (1.58)
		Overall	4,068	14.78 (0.99)	15.00 (1.00)
	Past Month	Yes	2,522	8.35 (0.93)	8.34 (0.91)
		No	1,546	9.66 (1.27)	9.61 (1.25)
		Overall	4,068	8.87 (0.76)	8.84 (0.74)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.5 Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

Drug	Parent Talked about Dangers with Child	2005		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	Yes	2,522	23.89 (1.52)	24.05 (1.52)
	No	1,546	27.24 (1.86)	27.20 (1.88)
	Overall	4,068	25.22 (1.16)	25.29 (1.17)
Past Year	Yes	2,522	18.58 (1.41)	18.64 (1.40)
	No	1,546	20.10 (1.65)	19.97 (1.65)
	Overall	4,068	19.18 (1.05)	19.17 (1.05)
Past Month	Yes	2,522	9.59 (1.10)	9.61 (1.09)
	No	1,546	11.43 (1.43)	11.35 (1.43)
	Overall	4,068	10.32 (0.87)	10.30 (0.86)
Marijuana				
Lifetime	Yes	2,522	14.50 (1.25)	14.57 (1.25)
	No	1,546	16.98 (1.49)	16.84 (1.49)
	Overall	4,068	15.48 (0.97)	15.47 (0.97)
Past Year	Yes	2,522	12.25 (1.21)	12.23 (1.21)
	No	1,546	14.03 (1.42)	13.97 (1.43)
	Overall	4,068	12.96 (0.92)	12.91 (0.93)
Past Month	Yes	2,522	6.33 (0.93)	6.25 (0.91)
	No	1,546	6.52 (1.03)	6.51 (1.04)
	Overall	4,068	6.40 (0.69)	6.35 (0.69)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.6a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Children about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

Drug	Mother Talked about Dangers with Child	2005 *n*	2005 Baseline[1]	2005 Final[2]
Alcohol				
Lifetime	0 times	140	30.26 (5.22)	29.33 (5.09)
	1-2 times	277	33.26 (4.49)	33.56 (4.46)
	A few times	613	38.26 (3.05)	38.44 (3.07)
	Many times	1,321	39.38 (2.13)	39.32 (2.15)
	Overall	2,351	37.82 (1.59)	37.82 (1.58)
Past Year	0 times	140	19.66 (4.29)	18.83 (4.09)
	1-2 times	277	26.14 (4.19)	26.33 (4.16)
	A few times	613	33.23 (3.03)	33.58 (3.04)
	Many times	1,321	31.64 (2.06)	32.07 (2.08)
	Overall	2,351	30.72 (1.49)	31.05 (1.50)
Past Month	0 times	140	9.54 (3.36)	9.34 (3.26)
	1-2 times	277	7.19 (1.77)	7.37 (1.82)
	A few times	613	17.38 (2.53)	17.44 (2.53)
	Many times	1,321	16.34 (1.56)	16.75 (1.61)
	Overall	2,351	15.15 (1.13)	15.43 (1.15)
Cigarettes				
Lifetime	0 times	140	12.93 (3.52)	12.42 (3.43)
	1-2 times	277	20.44 (3.92)	21.21 (4.00)
	A few times	613	20.44 (2.88)	20.91 (2.92)
	Many times	1,321	24.56 (1.85)	25.20 (1.88)
	Overall	2,351	22.27 (1.42)	22.79 (1.44)
Past Year	0 times	140	6.93 (2.61)	7.04 (2.75)
	1-2 times	277	11.01 (2.96)	11.65 (3.10)
	A few times	613	14.22 (2.53)	14.61 (2.58)
	Many times	1,321	17.54 (1.73)	18.11 (1.78)
	Overall	2,351	15.24 (1.26)	15.74 (1.29)
Past Month	0 times	140	4.00 (1.85)	3.60 (1.67)
	1-2 times	277	6.66 (2.50)	6.82 (2.55)
	A few times	613	8.06 (1.59)	8.13 (1.57)
	Many times	1,321	12.03 (1.51)	12.04 (1.50)
	Overall	2,351	9.84 (1.02)	9.85 (1.01)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.6b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

Drug	Father Talked about Dangers with Child	2005		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	0 times	181	34.31 (6.14)	34.44 (6.24)
	1-2 times	261	29.36 (4.48)	28.89 (4.46)
	A few times	508	37.57 (3.48)	38.09 (3.58)
	Many times	580	37.85 (3.05)	37.88 (3.09)
	Overall	1,530	35.75 (1.99)	35.86 (2.01)
Past Year	0 times	181	25.58 (5.38)	25.63 (5.39)
	1-2 times	261	19.92 (3.57)	19.60 (3.54)
	A few times	508	30.19 (3.32)	29.99 (3.34)
	Many times	580	27.35 (2.54)	27.00 (2.55)
	Overall	1,530	26.60 (1.72)	26.37 (1.71)
Past Month	0 times	181	10.75 (3.70)	10.20 (3.53)
	1-2 times	261	8.46 (2.53)	8.16 (2.40)
	A few times	508	12.79 (2.36)	12.86 (2.36)
	Many times	580	13.63 (2.01)	13.94 (2.10)
	Overall	1,530	12.05 (1.20)	12.07 (1.21)
Cigarettes				
Lifetime	0 times	181	15.49 (3.48)	16.43 (3.78)
	1-2 times	261	13.07 (2.57)	12.54 (2.46)
	A few times	508	18.43 (2.60)	18.34 (2.59)
	Many times	580	23.38 (2.44)	23.12 (2.43)
	Overall	1,530	18.97 (1.46)	18.90 (1.46)
Past Year	0 times	181	11.38 (3.03)	12.18 (3.37)
	1-2 times	261	9.21 (2.23)	8.73 (2.10)
	A few times	508	14.53 (2.40)	14.34 (2.37)
	Many times	580	15.03 (2.07)	14.85 (2.07)
	Overall	1,530	13.32 (1.27)	13.24 (1.26)
Past Month	0 times	181	6.02 (2.14)	6.09 (2.16)
	1-2 times	261	4.66 (1.48)	4.57 (1.49)
	A few times	508	6.03 (1.39)	6.05 (1.38)
	Many times	580	8.85 (1.64)	8.74 (1.66)
	Overall	1,530	6.88 (0.84)	6.84 (0.85)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.7a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

		2005		
Drug	Mother Talked about Dangers with Child	*n*	Baseline[1]	Final[2]
Any Illicit				
Lifetime	0 times	140	21.44 (4.43)	20.96 (4.34)
	1-2 times	277	22.39 (3.80)	22.37 (3.73)
	A few times	613	29.61 (3.05)	30.14 (3.06)
	Many times	1,321	26.94 (1.92)	26.82 (1.91)
	Overall	2,351	26.81 (1.48)	26.89 (1.47)
Past Year	0 times	140	17.29 (4.23)	16.90 (4.13)
	1-2 times	277	17.58 (3.47)	17.59 (3.41)
	A few times	613	22.97 (2.89)	23.20 (2.89)
	Many times	1,321	20.59 (1.80)	20.54 (1.80)
	Overall	2,351	20.69 (1.36)	20.73 (1.36)
Past Month	0 times	140	10.48 (3.41)	9.71 (3.17)
	1-2 times	277	6.18 (2.31)	6.15 (2.20)
	A few times	613	10.87 (1.99)	11.04 (1.97)
	Many times	1,321	12.99 (1.55)	12.92 (1.55)
	Overall	2,351	11.46 (1.05)	11.43 (1.05)
Marijuana				
Lifetime	0 times	140	10.13 (3.21)	9.45 (3.07)
	1-2 times	277	12.11 (2.84)	12.61 (2.98)
	A few times	613	19.11 (2.78)	19.62 (2.80)
	Many times	1,321	18.92 (1.78)	18.73 (1.77)
	Overall	2,351	17.66 (1.33)	17.73 (1.33)
Past Year	0 times	140	8.48 (3.07)	7.87 (2.93)
	1-2 times	277	9.90 (2.67)	10.43 (2.83)
	A few times	613	16.98 (2.76)	17.37 (2.79)
	Many times	1,321	15.59 (1.67)	15.43 (1.67)
	Overall	2,351	14.88 (1.25)	14.95 (1.26)
Past Month	0 times	140	3.16 (1.63)	2.77 (1.44)
	1-2 times	277	2.82 (1.23)	3.29 (1.53)
	A few times	613	6.11 (1.65)	6.32 (1.64)
	Many times	1,321	9.64 (1.37)	9.54 (1.38)
	Overall	2,351	7.50 (0.90)	7.52 (0.90)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 8.7b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2005

Drug	Father Talked about Dangers with Child	2005 *n*	2005 Baseline[1]	2005 Final[2]
Any Illicit				
Lifetime	0 times	181	14.43 (3.30)	14.46 (3.32)
	1-2 times	261	19.13 (4.47)	18.91 (4.56)
	A few times	508	22.93 (3.13)	23.16 (3.22)
	Many times	580	25.13 (2.78)	25.08 (2.84)
	Overall	1,530	21.88 (1.63)	21.91 (1.68)
Past Year	0 times	181	12.60 (3.16)	12.55 (3.16)
	1-2 times	261	12.31 (4.10)	12.42 (4.23)
	A few times	508	19.12 (3.10)	19.40 (3.20)
	Many times	580	16.68 (2.07)	16.39 (2.06)
	Overall	1,530	16.03 (1.46)	16.02 (1.50)
Past Month	0 times	181	5.93 (2.36)	5.87 (2.35)
	1-2 times	261	4.96 (3.92)	5.06 (4.06)
	A few times	508	8.62 (1.97)	8.71 (1.98)
	Many times	580	8.73 (1.60)	8.95 (1.68)
	Overall	1,530	7.63 (1.13)	7.75 (1.16)
Marijuana				
Lifetime	0 times	181	9.95 (2.75)	9.82 (2.71)
	1-2 times	261	5.94 (1.42)	5.78 (1.41)
	A few times	508	12.08 (2.53)	12.63 (2.73)
	Many times	580	14.82 (1.99)	14.23 (1.94)
	Overall	1,530	11.74 (1.16)	11.64 (1.19)
Past Year	0 times	181	8.07 (2.45)	7.87 (2.39)
	1-2 times	261	4.71 (1.29)	4.67 (1.31)
	A few times	508	9.37 (2.45)	9.86 (2.67)
	Many times	580	12.17 (1.84)	11.67 (1.79)
	Overall	1,530	9.44 (1.08)	9.36 (1.12)
Past Month	0 times	181	3.17 (1.53)	3.07 (1.48)
	1-2 times	261	0.82 (0.56)	0.75 (0.53)
	A few times	508	4.83 (1.49)	4.88 (1.49)
	Many times	580	5.62 (1.31)	5.61 (1.34)
	Overall	1,530	4.18 (0.71)	4.17 (0.72)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

References

Aldworth, J., Ault, K., Barnett-Walker, K., Carpenter, L., Copello, E., Frechtel, P., Licata, A., Liu, B., Martin, P., & Tillman, P. (2007, February). Imputation report [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.177.007). Research Triangle Park, NC: RTI International. [To be available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

Chen, P., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2007, January). Person-level sampling weight calibration [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009). Research Triangle Park, NC: RTI International. [To be available as a PDF at http://www.oas.samhsa.gov/nhsda/methods.cfm#2k5]

Chen, P., Penne, M. A., & Singh, A. C. (2000). Experience with the generalized exponential model (GEM) for weight calibration for NHSDA. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 604-607). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights. In *Proceedings of the 2001 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the 1980 American Statistical Association, Survey Research Methods Section, Houston, TX* (pp. 721-726). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376-382.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

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

Hajek, J. (1971). Comment on D. Basu's paper, "An essay on the logical foundations of survey sampling, part one" [pp. 203-234]. In V. P. Godambe, & D. A. Sprott (Eds.), *Foundations of statistical inference: A symposium* (p. 236). Toronto, Ontario, & Montreal, Quebec, Canada: Holt, Rinehart and Winston of Canada.

Iannacchione, V. (1982). Weighted sequential hot deck imputation macros. In *Proceedings of the Seventh Annual SAS Users Group International Conference* (pp. 759-763). Cary, NC: SAS Corporation.

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

Morton, K. B., Chromy, J. R., Hunter, S. R., & Martin, P. C. (2006, February). Sample design report [2005]. In *2005 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Deliverable No. 8, RTI/0209009). Research Triangle Park, NC: RTI International. [Available as a PDF at http://www.oas.samhsa.gov/nsduh/methods.cfm#2k5]

Penne, M., Chen, P., & Singh, A.C. (2001) Person-pair Sampling Weight Calibration using the Generalized Exponential Model for The National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

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

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). Boca Raton, FL: Chapman and Hall/CRC.

Singh, A., Grau, E., & Folsom, R., Jr. (2001). Predictive mean neighborhood imputation with application to the person-pair data of the National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

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

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

Appendix A: Technical Details about the Generalized Exponential Model

Appendix A: Technical Details about the Generalized Exponential Model

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

$$\Delta(w,d) = \sum_{k \in s} \frac{d_k}{A_k} \cdot \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - a_k} \right\}, \qquad \text{(A.1.1)}$$

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

$$\sum_{k \in s} x_k d_k a_k = T_x. \qquad \text{(A.1.2)}$$

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

$$a_k(\lambda) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}{(u_k - c_k) + (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}. \qquad \text{(A.1.3)}$$

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 A.1.3 that

$$\ell_k < a_k < u_k, \; k = 1, \ldots, n. \qquad \text{(A.1.4)}$$

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

$$\Delta(w,d) = \sum_{k \in s} d_k a_k \log a_k - \sum_{k \in s} d_k (a_k - 1) \qquad \text{(A.1.5)}$$

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 (A.1.3) 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 nonextreme weight subsample, and s^{**}, the extreme weight subsample. For nonextreme weights, the following are set: $\ell_2 = 1$, $c_2 = \rho^{-1}$, $u_2 = u > \rho^{-1}$, where ρ is the overall response propensity. For extreme weights with high weights, $\ell_k = \ell m_k$, $c_k = \rho^{-1} m_k$, and $u_k = u_1 m_k$, 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 nonextreme weights, $\ell_k = \ell_2$, $c_k = c_2 = 1$, and $u_k = u_2$; for high extreme weights, $\ell_2 = \ell_1 m_k$, $c_k = m_k$, and $u_k = u_1 m_k$; and, similarly for low extreme weights, $\ell_k = \ell_3 m_k$, $c_k = m_k$, and $u_k = u_3 m_k$. 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 surveys, constrained exponential models were used for poststratification, and scaled constrained exponential models were used 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 could not be derived from the U.S. Bureau of the Census directly. Estimates of the number of households and person pairs were not available at the domains that we wanted to control, and person pair population estimates were 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 could be poststratified to census population estimates. The calibrated SDU weights then could 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 were calibrated to a second set of controls derived from the questionnaire, called household-level person counts. These controls were applied to pairs that were 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 younger than the age of 18 living in the household.

B.1 Derivation of QDU-Level Poststratification Controls

The derivation of QDU-level poststratification controls was not directly possible. Instead, it had to be based on work done for the person-level calibration. At the person level, weights were calibrated to the control totals that we wished to reach. These weights then were 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 or older were provided by the Population Estimates Branch of the U.S. Bureau of the Census. We received 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 (11).

The breakdown received from the census did not match the levels of the domains that we wanted to control. To account for this, we collapsed levels. From this altered data, we created datasets with model group-specific control totals. Observations in these datasets corresponded to a breakdown by quarter (4), Hispanicity (2), race (5), sex (2), age (11), and number of States[27] in the model group (number of States varied according to which census region was represented in the model group).

B.1.1.2 Adjusting SDU Data to the Control Totals

In the person-level weighting, the SDU weights were poststratified to meet control totals based on the population estimates received from the census. For NSDUH weighting, GEM was utilized to calibrate sample weights to multiple control totals. In doing so, each SDU received an adjustment factor, which, when multiplied by the initial weight, produced a final weight. The sum of all final weights corresponded to the civilian, noninstitutionalized population estimate for ages 12 or older, and the sum of all final weights in a domain corresponded to the control total for that domain. Note that there were a number of controls being calibrated to for each SDU, depending upon the domains to which the SDU belonged. The adjusted SDU weight reflected the civilian, noninstitutionalized population estimates for ages 12 or older and could 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 were no controls for QDU-level poststratification available directly, we used the adjusted SDU weights. For these weights to be applicable at the QDU level, the SDU-level data had to be restructured by sorting and summing over the domains to be used in the QDU-level calibration. This provided a dataset where the summed weight, which still added up to the proper population, was available for every domain to be utilized in the QDU calibration and thus could be used as a control total.

[27] 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 was adjusted via calibration in GEM of sample weights to multiple control totals. Each QDU received an adjustment factor, similar to that described for the SDU weight in B.1.1.2. The controls utilized in this calibration were based on the SDU weight as described in B.1.2.1 above. The adjusted weight was representative of the civilian, noninstitutionalized population estimates for ages 12 or 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 Household-Level Person Counts

Analogous to the QDU weights, some of the person pair controls were 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 one 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 dwelling units) 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 for details on the household-level control totals, which are referred to as household-level person counts.

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

Like the SDU- and QDU-level data, the person pair-level data was adjusted via GEM. The use of two different types of controls required a minor modification to the GEM macro so that both sets of controls might be addressed simultaneously. Similar to the SDU- and QDU-level poststratification steps, each pair received an adjustment factor, which, when multiplied by the initial weight, produced a final weight. The sum of all final weights corresponded to the civilian, noninstitutionalized population estimate for ages 12 or older, and the sum of all final weights in a domain corresponded 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 2005, 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 C.1a through C.4b. 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 led to situations whereby inclusion of all originally proposed levels of covariates in the model was not possible. The text and exhibits in Sections C.1 through C.4 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, and for the list of realized final model covariates, see Exhibits C.1.1 through C.4.3. 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 were present in all stages of modeling, and the initial set of variables was the same across model groups but may change over stages of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment. Exhibits C.1.1 through C.4.3 provide lists of the proposed and realized covariates.

To help understand what effects were controlled for at each stage of the modeling, it was 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[b]
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1]
Gender[b]
1: Male, 2: Female[1]
Group Quarter Indicator[a]
1: College Dorm, 2: Other Group Quarter, 3: Nongroup Quarter[1]
Hispanicity[b]
1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]
Household Size[b]
Continuous variable count of individuals rostered with DU
Household Type (Ages of Persons Rostered within DU)[a]
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+[1]
Percentage of Owner-Occupied Dwelling Units in Segment (% Owner)[a]
1: 50-100%,[1] 2: 10->50%, 3: 0->10%
Percentage of Segments That Are Black or African American (% Black)[a]
1: 50-100%, 2: 10->50%, 3: 0->10%[1]
Percentage of Segments That Are Hispanic or Latino (% Hispanic)[a]
1: 50-100%, 2: 10->50%, 3: 0->10%[1]
Population Density[a]
1: MSA 1,000,000 or more, 2: MSA less than 1,000,000, 3: Non-MSA urban, 4: Non-MSA rural[1]
Quarter[a,b]
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]
Race (3 Levels)[b]
1: white[1], 2: black or African American, 3: other
Race (5 Levels)[b]
1: white,[1] 2: black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: multirace
Race/Ethnicity of Householder[a]
1: Hispanic or Latino white,[1] 2: Hispanic or Latino black or African American, 3: Hispanic or Latino others, 4: Non-Hispanic or Latino white, 5: Non-Hispanic or Latino black or African American, 6: Non-Hispanic or Latino others
Relation to Householder[a]
1: Householder or Spouse, 2: Child, 3: Other Relative, 4: Nonrelative[1]
Segment-Combined Median Rent and Housing Value (Rent/Housing)[a,2]
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[a,b,3]	
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[a,3]	
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]

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.

[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.

[a] Binary variable.

[b] 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. The factor effect represents the effects of levels considered for one factor, two factors, 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 or all of the factor effects in the reference set.

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

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

Hier. One or more of the effects in a higher order interaction is 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 wild card character to indicate all levels of the factor for that effect.

*Note: The above glossary is 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 2004 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 or African American, 3: other

Note that the superscript numbers indicate 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 or African American	Asian	American Indian or Alaska Native

Indicates the reference-level set.

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

State*Race (4 Levels)	white	black or African American	Asian	American Indian or 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 or African American	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 or African American</th><th>Asian</th><th>American Indian or 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 or African American	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 2005 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/Ethnicity of Householder	Y		6	5
Rent/housing Value	Y		5	4
Segment % Black or African American	Y		3	2
Segment % Hispanic or Latino	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 or African American x % Owner	Y		3 x 3	4
% Black or African American 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 C.1: Model Group 1: Northeast

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

Table C.1a 2005 QDU Weight GEM Modeling Summary (Model Group 1: Northeast)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	1.72	1.63	0.32	2.911	243	(0.38, 2.70)	(0.38, 2.70)
	1.66	2.79	0.62	2.928	242	(0.20, 2.95)	(0.20, 2.95)
						(0.40, 1.45)	(0.40, 1.45)
res.qdu.nr	1.59	2.71	0.63	2.985	243	(1.00, 3.00)	(1.00, 3.00)
	1.50	2.72	0.77	3.451	243	(1.00, 5.00)	(1.00, 5.00)
						(1.00, 1.42)	(1.00, 1.41)
res.qdu.ps	1.50	2.72	0.77	3.451	243	(0.95, 2.00)	(0.97, 2.00)
	1.79	2.94	0.59	3.455	242	(0.85, 2.00)	(0.86, 1.35)
						(0.95, 1.08)	(0.95, 1.05)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.
[2] Unequal weighting effect (UWE) 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 C.1b 2005 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	**duwght11**	**1-11**	**duwght12**	**1-12**	**duwght13**	**1-13**	**duwght14**	**1-14**
Minimum	22	1.00	22	0.20	8	0.57	9	0.58	8
1%	76	1.00	96	0.51	90	1.00	98	0.90	96
5%	150	1.00	186	0.71	175	1.00	187	0.96	187
10%	217	1.00	284	0.80	264	1.03	283	0.98	283
25%	412	1.00	593	0.91	572	1.08	612	0.99	611
Median	669	1.07	885	1.01	883	1.15	979	1.00	979
75%	953	3.18	1,835	1.10	1,890	1.27	2,104	1.01	2,104
90%	1,309	5.95	4,396	1.22	4,596	1.42	5,469	1.02	5,441
95%	1,650	8.38	6,788	1.35	7,286	1.54	9,288	1.03	9,285
99%	2,257	13.04	12,297	1.72	12,407	2.05	16,974	1.08	17,088
Maximum	6,971	14.71	25,305	3.96	28,144	5.00	41,341	1.98	37,820
n	11,599	-	11,599	-	11,599	-	9,617	-	9,617
Mean	743	2.41	1,768	1.02	1,820	1.21	2,195	1.00	2,195
Max/Mean	9.38	-	14.32	-	15.47	-	18.84	-	17.23

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 1 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native and Asian in Connecticut were collapsed together due to convergence problems. All three-factor effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All main effects were maintained in full. Among two-factor effects, one level of the segment percent Hispanic or Latino by segment rent/housing variable was collapsed due to zero counts, and the race levels of American Indian or Alaska Native, Asian, and Multiple Race in Maine were collapsed due to convergence problems. All three-factor effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Poststratification

All main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native and Asian in Connecticut were collapsed together due to convergence problems. All three-factor effects were maintained in full.

Exhibit C.1.1 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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	31	Coll. (1,3) & (1,4); conv.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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		**243**	**242**	

Exhibit C.1.2 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**133**	
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 or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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		**243**	**243**	

Exhibit C.1.3 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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	31	Coll. (1,3) & (1,4); conv.
State x Gender	9 x 2	8	8	All levels present.
State x Hispanicity	9 x 2	8	8	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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		**243**	**242**	

Appendix C.2: Model Group 2: Midwest

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

Table C.2a 2005 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	2.30	2.14	0.35	2.621	300	(0.25, 2.80)	(0.25, 2.80)
	1.28	1.30	0.24	2.587	300	(0.29, 2.80)	(0.30, 2.72)
						(0.55, 1.77)	(0.55, 1.74)
res.qdu.nr	1.45	1.47	0.26	2.612	300	(1.00, 2.70)	(1.00, 2.70)
	0.86	1.20	0.29	2.915	292	(1.00, 4.28)	(1.00, 4.24)
						(1.00, 3.93)	(1.00, 3.93)
res.qdu.ps	0.86	1.20	0.29	2.915	300	(0.35, 2.25)	(0.35, 2.25)
	0.88	1.11	0.15	2.910	287	(0.21, 2.25)	(0.21, 1.81)
						(0.30, 1.15)	(0.30, 1.14)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect (UWE) 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 C.2b 2005 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	**duwght11**	**1-11**	**duwght12**	**1-12**	**duwght13**	**1-13**	**duwght14**	**1-14**
Minimum	23	1.00	23	0.08	15	0.47	15	0.21	17
1%	109	1.00	115	0.61	114	1.00	119	0.89	118
5%	156	1.00	208	0.81	204	1.01	219	0.97	215
10%	266	1.00	387	0.86	371	1.04	391	0.99	389
25%	519	1.00	590	0.93	585	1.09	648	0.99	649
Median	634	1.06	790	0.99	802	1.16	914	1.00	915
75%	802	2.92	1,821	1.06	1,786	1.26	2,028	1.01	2,030
90%	1,300	5.49	3,920	1.16	3,922	1.40	4,981	1.01	4,978
95%	1,513	7.43	5,596	1.25	5,647	1.48	7,233	1.03	7,235
99%	1,925	11.75	10,340	1.56	9,935	1.93	13,143	1.12	13,140
Maximum	6,105	13.90	35,086	4.83	29,386	6.51	43,712	1.88	35,881
n	15,996	-	15,996	-	15,996	-	13,342	-	13,342
Mean	715	2.29	1,622	1.01	1,620	1.20	1,942	1.00	1,942
Max/Mean	8.54	-	21.63	-	18.14	-	22.51	-	18.48

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 2 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All 300 proposed effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native, Asian, and Multiple Race were collapsed together in Indiana due to convergence problems. Among three-factor effects, the States of Illinois, Michigan, and Ohio each combined the race levels of black or African American with Other when crossed by Hispanicity. Ohio age levels were combined into the 18-or-older group in the three-factor interaction of State/Region, age, and Hispanicity.

Respondent Questionnaire Dwelling Unit-Level Poststratification

The race of householder levels of Hispanic or Latino black or African American and Hispanic or Latino others were combined. All other main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native and Asian were collapsed together in all States due to convergence problems. Further, in North Dakota, all race levels of black or African American and Multiple Race were combined. All three-factor effects were maintained in full.

Exhibit C.2.1 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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 or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American x Rent/Housing	3 x 5	8	8	All levels present.
% Hispanicity or Latino x % Owner-Occupied	3 x 3	4	4	All levels present.
% Hispanicity or Latino 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 C.2.2 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**161**	
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	42	Coll. (3,3), (3,4) & (3.5); conv.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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**	**65**	
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	9	Coll. (10,2), (10,3), (10,4) & (10,5); conv.
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	3	Coll. (1,2,1) & (1,3,1). Repeat for State/Regions (2) and (3).
State/Region x Race (3 Levels) x Gender	4 x 3 x 2	6	6	All levels present.
Total		**300**	**292**	

Exhibit C.2.3 Covariates for 2005 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**65**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	4	Coll. (2) & (3); conv.
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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**151**	
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	32	Coll. (2,3) & (2,4); conv. Repeat for all States. Coll. (9,2) & (9,5); conv.
State x Gender	12 x 2	11	11	All levels present.
State x Hispanicity	12 x 2	11	11	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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**	**287**	

Appendix C.3: 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 C.3a 2005 QDU Weight GEM Modeling Summary (Model Group 3: South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	1.56	1.40	0.26	2.505	339	(0.45, 2.25)	(0.47, 2.25)
	1.16	1.54	0.35	2.537	338	(0.20, 4.22)	(0.20, 4.22)
						(0.99, 2.49)	(0.99, 2.49)
res.qdu.nr	1.34	2.31	0.49	2.562	339	(1.00, 2.05)	(1.00, 2.05)
	1.14	2.27	0.40	2.891	335	(1.00, 4.44)	(1.00, 4.44)
						(1.05, 1.50)	(1.05, 1.48)
res.qdu.ps	1.14	2.27	0.40	2.891	339	(0.85, 1.60)	(0.97, 1.60)
	1.04	1.94	0.19	2.887	337	(0.20, 1.60)	(0.25, 1.31)
						(0.99, 1.15)	(0.99, 1.15)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.
[2] Unequal weighting effect (UWE) 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 C.3b 2005 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	**duwght11**	**1-11**	**duwght12**	**1-12**	**duwght13**	**1-13**	**duwght14**	**1-14**
Minimum	36	1.00	36	0.20	26	0.41	27	0.25	24
1%	71	1.00	93	0.60	84	1.00	95	0.92	94
5%	136	1.00	198	0.79	197	1.01	220	0.98	219
10%	268	1.00	406	0.84	398	1.03	447	0.99	448
25%	672	1.00	821	0.92	801	1.07	887	1.00	887
Median	977	1.12	1,294	0.99	1,291	1.14	1,429	1.00	1,430
75%	1,323	3.15	2,614	1.08	2,608	1.26	2,749	1.00	2,753
90%	1,704	6.03	5,741	1.17	5,651	1.41	6,948	1.01	6,930
95%	2,063	7.43	8,175	1.25	8,116	1.51	10,498	1.02	10,494
99%	2,881	11.71	14,300	1.62	14,414	1.88	19,348	1.07	19,433
Maximum	9,584	16.61	31,319	7.12	39,585	4.64	43,695	1.73	44,318
n	17,579	-	17,579	-	17,579	-	14,744	-	14,744
Mean	1,025	2.37	2,340	1.01	2,342	1.19	2,792	1.00	2,792
Max/Mean	9.35	-	13.38	-	16.91	-	15.65	-	15.87

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 3 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native and Asian were collapsed together in West Virginia due to convergence problems. All three-factor effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All main effects were maintained in full. Among two-factor effects, the District of Columbia, South Carolina, and West Virginia collapsed the race levels of American Indian or Alaska Native and Asian due to convergence problems. In Louisiana, the race levels of American Indian or Alaska Native and Multiple Race were collapsed together. All three-factor effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Poststratification

All main effects were maintained in full. Among two-factor effects, the race levels of American Indian or Alaska Native and Asian were collapsed together in the District of Columbia and West Virginia due to convergence problems. All three-factor effects were maintained in full.

Exhibit C.3.1 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**212**	
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	63	Coll. (17,3) & (17,4); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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**	**338**	

Exhibit C.3.2 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**209**	
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	60	Coll. (4,3) & (4,4); conv. Repeat for States (13) and (17). Coll. (8,3) & (8,5); conv.
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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**	**335**	

Exhibit C.3.3 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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. (4,3) & (4,4); conv. Repeat for State (17).
State x Gender	17 x 2	16	16	All levels present.
State x Hispanicity	17 x 2	16	16	All levels present.
% Black or African American x % Owner-Occupied	3 x 3	4	4	All levels present.
% Black or African American 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 C.4: Model Group 4: West

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

Table C.4a 2005 QDU Weight GEM Modeling Summary (Model Group 4: West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	1.70	1.25	0.25	2.959	270	(0.55, 1.80)	(0.55, 1.80)
	1.23	1.85	0.30	2.946	265	(0.26, 2.91)	(0.26, 2.91)
						(0.99, 4.50)	(0.99, 4.50)
res.qdu.nr	1.55	2.47	0.41	2.942	270	(1.00, 1.50)	(1.00, 1.50)
	1.21	2.28	0.33	3.363	265	(1.00, 3.07)	(1.00, 3.07)
						(1.20, 2.31)	(1.20, 2.31)
res.qdu.ps	1.21	2.28	0.33	3.363	270	(0.90, 1.25)	(0.98, 1.25)
	1.30	2.13	0.17	3.366	265	(0.70, 1.50)	(0.76, 1.44)
						(0.99, 1.50)	(0.99, 1.09)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.
[2] Unequal weighting effect (UWE) 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 C.4b 2005 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13	duwght14	1-14
Minimum	17	1.00	17	0.24	17	0.53	20	0.64	20
1%	77	1.00	84	0.51	76	1.00	78	0.92	78
5%	111	1.00	129	0.74	127	1.00	135	0.99	134
10%	140	1.00	181	0.80	175	1.02	191	0.99	190
25%	275	1.00	403	0.89	393	1.07	428	1.00	429
Median	677	1.08	1,169	0.98	1,127	1.13	1,237	1.00	1,237
75%	1,631	2.93	2,205	1.08	2,236	1.24	2,510	1.00	2,510
90%	2,064	5.14	5,408	1.19	5,151	1.40	6,051	1.01	6,033
95%	2,256	7.34	7,803	1.28	7,810	1.52	9,764	1.02	9,748
99%	2,752	11.82	14,442	1.52	13,583	1.93	18,248	1.05	18,357
Maximum	6,540	14.50	32,377	8.07	27,246	4.46	43,698	1.44	42,835
n	12,069	-	12,069	-	12,069	-	10,190	-	10,190
Mean	949	2.27	2,072	0.99	2,027	1.18	2,400	1.00	2,400
Max/Mean	6.89	-	15.63	-	13.44	-	18.20	-	17.85

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 4 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All main effects were maintained in full. Among two-factor effects, all race levels of the segment percent black or African American by segment rent/housing variable were collapsed due to singularity problems. The levels of percent owner-occupied dwelling units in segment also were combined for the race level of black or African American to produce a category of less than 50 percent. All three-factor effects were maintained in full.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

This model was identical to the selected questionnaire dwelling unit-level poststratification step.

Respondent Questionnaire Dwelling Unit-Level Poststratification

This model was identical to the selected questionnaire dwelling unit-level poststratification step and the respondent questionnaire dwelling unit-level nonresponse step.

Exhibit C.4.1 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**168**	
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 or African American x % Owner-Occupied	3 x 3	4	3	Coll. (3,1) & (3,2); conv.
% Black or African American x Rent/Housing	3 x 5	8	4	Coll. (1,1) & (2,1); sing. Repeat for all levels of Rent/Housing.
% 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**	**265**	

Exhibit C.4.2 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**168**	
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 or African American x % Owner-Occupied	3 x 3	4	3	Coll. (3,1) & (3,2); conv.
% Black or African American x Rent/Housing	3 x 5	8	4	Coll. (1,1) & (2,1); sing. Repeat for all levels of Rent/Housing.
% 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**	**265**	

Exhibit C.4.3 Covariates for 2005 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/Ethnicity 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 or African American	3	2	2	All levels present.
% Hispanic or Latino	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**	**168**	
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 or African American x % Owner-Occupied	3 x 3	4	3	Coll. (3,1) & (3,2); conv.
% Black or African American x Rent/Housing	3 x 5	8	4	Coll. (1,1) & (2,1); sing. Repeat for all levels of Rent/Housing.
% 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**	**265**	

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

Table D.1 2005 NSDUH QDU-Level Response Rates

Domain	Selected QDU	Respondent QDU	% Interview Response Rate[1]
Total	57,243	47,893	78.16
Census Region			
Northeast	11,599	9,617	75.98
South	17,579	14,744	78.54
Midwest	15,996	13,342	78.24
West	12,069	10,190	79.33
Quarter			
Quarter 1	13,629	11,424	78.23
Quarter 2	15,342	12,800	77.96
Quarter 3	14,157	11,900	78.30
Quarter 4	14,115	11,769	78.16
Household Type			
12-17, 18-25, 26+	5,197	4,634	88.51
12-17, 18-25	132	111	82.21
12-17, 26+	16,751	14,757	87.88
18-25, 26+	11,310	9,477	83.94
12-17	35	33	96.22
18-25	6,735	5,964	87.96
26+	17,083	12,917	73.93
Race/Ethnicity of Householder			
Hispanic or Latino White	6,467	5,485	80.82
Hispanic or Latino Black or African American	117	105	86.04
Hispanic or Latino Other	442	388	78.08
Non-Hispanic or Latino White	39,900	33,101	77.28
Non-Hispanic or Latino Black or African American	6,632	5,772	82.30
Non-Hispanic or Latino Other	3,685	3,042	75.33
% Hispanic or Latino in Segment			
50-100%	3,983	3,370	80.28
10-50%	10,169	8,517	79.33
<10%	43,091	36,006	77.62
% Black or African American in Segment			
50-100%	4,430	3,820	80.58
10-50%	8,744	7,414	79.76
<10%	44,069	36,659	77.55
% Owner-Occupied DUs in Segment			
50-100%	43,173	35,934	77.76
10-50%	10,916	9,235	78.96
<10%	3,154	2,724	80.80
Combined Median Rent/Housing Value			
1st Quintile	9,237	7,971	80.89
2nd Quintile	12,537	10,700	79.58
3rd Quintile	12,717	10,655	78.57
4th Quintile	12,406	10,233	77.31
5th Quintile	10,346	8,334	75.43
Population Density			
Large MSA	24,515	20,074	76.72
Medium to Small MSA	27,647	23,434	79.48
Non-MSA, Urban	1,536	1,345	83.39
Non-MSA, Rural	3,545	3,040	79.85
Group Quarters			
Group	854	822	92.86
Nongroup	56,389	47,071	78.04
Household Size			
One	7,114	5,673	74.42
Two	21,333	17,353	76.76
Three	15,900	13,581	83.13
Four or More	12,896	11,286	86.02

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] The weight used for calculating the response rate includes SDU- and QDU-level design weights, SDU nonresponse and poststratification adjustments, and selected QDU poststratification adjustment. This weight is the product of YR05WT1*...*YR05WT9*DU05WT10*DU05WT11.

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

Table E.1 2005 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)			Before sel.qdu.ps[1] (SDUWT*DU05WT10)			After sel.qdu.ps[1] (SDUWT*DU05WT10*DU05WT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	57,243	2.33	3.42	0.69	1.83	1.58	0.29	1.31	1.78	0.36
Census Region										
Northeast	11,599	2.31	3.84	0.83	1.72	1.63	0.32	1.66	2.79	0.62
South	17,579	1.97	3.27	0.63	1.56	1.40	0.26	1.16	1.54	0.35
Midwest	15,996	2.73	3.85	0.77	2.30	2.14	0.35	1.28	1.30	0.24
West	12,069	2.35	2.91	0.59	1.70	1.25	0.25	1.23	1.85	0.30
Quarter										
Quarter 1	13,629	2.61	3.56	0.68	1.97	1.59	0.27	1.42	1.83	0.39
Quarter 2	15,342	1.95	2.47	0.36	1.56	1.19	0.15	1.10	1.30	0.28
Quarter 3	14,157	2.42	4.06	0.91	1.90	1.81	0.39	1.28	2.08	0.44
Quarter 4	14,115	2.39	3.58	0.79	1.91	1.74	0.35	1.48	1.92	0.34
Household Type										
12-17, 18-25, 26+	5,197	2.25	3.23	0.72	2.25	3.23	0.72	2.04	4.40	1.00
12-17, 18-25	132	2.27	2.52	0.72	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	16,751	2.61	3.58	0.72	2.64	3.59	0.72	1.76	3.75	0.90
18-25, 26+	11,310	2.18	3.89	0.92	2.12	3.50	0.79	1.79	4.08	0.88
12-17	35	8.57	5.93	1.59	0.00	0.00	0.00	0.00	0.00	0.00
18-25	6,735	3.24	4.63	0.75	3.06	4.15	0.64	1.43	2.65	0.47
26+	17,083	1.82	2.53	0.45	0.24	0.58	0.07	0.30	0.81	0.13
Race/Ethnicity of Householder										
Hispanic or Latino White	6,467	2.61	2.77	0.49	2.13	1.44	0.20	1.30	1.43	0.35
Hispanic or Latino Black or African American	117	52.99	65.89	26.37	42.74	47.25	16.68	58.97	58.72	23.31
Hispanic or Latino Other	442	28.51	36.07	7.80	22.17	17.46	3.71	9.50	17.94	4.08
Non-Hispanic or Latino White	39,900	0.96	1.41	0.20	0.76	0.56	0.08	0.41	0.52	0.06
Non-Hispanic or Latino Black or African American	6,632	3.92	6.64	1.36	3.09	4.14	0.70	2.96	4.38	0.73
Non-Hispanic or Latino Other	3,685	9.04	11.90	2.32	6.87	5.94	1.18	5.35	7.94	1.55

Table E.1 2005 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	n	SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)			Before sel.qdu.ps[1] (SDUWT*DU05WT10)			After sel.qdu.ps[1] (SDUWT*DU05WT10*DU05WT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
% Hispanic or Latino in Segment										
50-100%	3,983	2.01	3.46	0.87	1.86	1.93	0.46	1.86	2.70	0.70
10-50%	10,169	2.97	4.08	0.84	2.22	1.62	0.32	1.75	2.48	0.53
<10%	43,091	2.21	3.21	0.62	1.73	1.53	0.26	1.16	1.49	0.28
% Black or African American in Segment										
50-100%	4,430	2.93	5.61	1.13	2.33	4.02	0.62	2.87	4.26	0.79
10-50%	8,744	2.80	4.32	0.98	2.24	2.12	0.44	1.89	3.10	0.73
<10%	44,069	2.18	2.96	0.57	1.70	1.22	0.22	1.04	1.23	0.24
% Owner-Occupied DUs in Segment										
50-100%	43,173	2.00	2.78	0.54	1.59	1.35	0.24	1.08	1.55	0.29
10-50%	10,916	3.28	5.10	1.11	2.51	2.20	0.44	1.91	2.25	0.54
<10%	3,154	3.65	5.45	0.98	2.73	2.47	0.47	2.41	3.27	0.72
Combined Median Rent/Housing Value										
1st Quintile	9,237	2.27	3.05	0.59	1.67	1.36	0.25	1.56	1.68	0.32
2nd Quintile	12,537	2.30	3.14	0.63	1.91	1.60	0.28	1.35	1.60	0.38
3rd Quintile	12,717	2.38	3.28	0.55	1.85	1.50	0.25	1.28	1.99	0.45
4th Quintile	12,406	2.42	3.68	0.93	1.93	1.71	0.39	1.08	1.53	0.31
5th Quintile	10,346	2.26	3.82	0.68	1.72	1.66	0.26	1.36	2.10	0.34
Population Density										
Large MSA[1]	24,515	2.18	3.62	0.77	1.77	1.81	0.34	1.55	2.26	0.43
Medium to Small MSA[1]	27,647	2.58	3.40	0.63	1.96	1.41	0.26	1.13	1.22	0.26
Non-MSA,[1] Urban	1,536	1.69	2.21	0.40	1.37	0.83	0.06	1.37	1.29	0.30
Non-MSA,[1] Rural	3,545	1.75	1.91	0.30	1.38	0.99	0.11	1.10	1.72	0.48
Group Quarters										
Group	854	3.63	6.30	0.75	2.93	3.72	0.36	3.16	5.13	0.65
Nongroup	56,389	2.31	3.38	0.69	1.81	1.56	0.29	1.28	1.76	0.36
Household Size										
One	7,114	1.90	3.03	0.64	1.14	0.75	0.11	0.69	0.66	0.09
Two	21,333	2.14	3.14	0.68	1.35	1.23	0.24	0.87	1.41	0.29
Three	15,900	2.39	3.26	0.59	2.27	2.68	0.45	1.58	3.11	0.65
Four or More	12,896	2.81	4.23	0.84	2.45	3.47	0.72	2.05	4.15	0.91

[1] DU = dwelling unit, MSA = metropolitan statistical area, PS = poststratification adjustment, QDU = questionnaire dwelling unit, SDU = screener dwelling unit, Sel = selected.
[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 nonextreme 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 2005 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors

Domain	n	Before res.qdu.nr[1] (SDUWT*DU05WT10*DU05WT11) % Unweighted	% Weighted[2]	% Outwinsor[3]	After res.qdu.nr[1] (SDUWT*DU05WT10*...*DU05WT12) % Unweighted	% Weighted[2]	% Outwinsor[3]	Final Weight: After res.qdu.ps[1] (SDUWT*DU05WT10*...*DU05WT13) % Unweighted	% Weighted[2]	% Outwinsor[3]
Total	47,893	1.47	2.22	0.45	1.15	2.11	0.43	1.20	1.98	0.25
Census Region										
Northeast	9,617	1.59	2.71	0.63	1.50	2.72	0.77	1.79	2.94	0.59
South	14,744	1.34	2.31	0.49	1.14	2.27	0.40	1.04	1.94	0.19
Midwest	13,342	1.45	1.47	0.26	0.86	1.20	0.29	0.88	1.11	0.15
West	10,190	1.55	2.47	0.41	1.21	2.28	0.33	1.30	2.13	0.17
Quarter										
Quarter 1	11,424	1.60	2.24	0.44	1.29	2.10	0.46	1.33	1.90	0.27
Quarter 2	12,800	1.19	1.54	0.32	1.08	1.92	0.43	1.21	1.95	0.24
Quarter 3	11,900	1.50	2.95	0.63	1.05	2.59	0.46	1.08	2.05	0.25
Quarter 4	11,769	1.61	2.15	0.39	1.19	1.83	0.38	1.19	2.01	0.24
Household Type										
12-17, 18-25, 26+	4,634	2.09	4.47	1.12	1.70	4.47	1.03	1.64	4.33	0.87
12-17, 18-25	111	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	14,757	1.84	3.87	0.93	1.45	3.85	1.03	1.48	3.57	0.60
18-25, 26+	9,477	1.81	3.85	0.90	1.69	4.68	1.10	1.79	4.52	0.65
12-17	33	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18-25	5,964	1.58	2.82	0.49	0.91	2.35	0.49	1.12	2.62	0.32
26+	12,917	0.53	1.34	0.20	0.33	1.21	0.17	0.33	1.07	0.08
Race/Ethnicity of Householder										
Hispanic or Latino White	5,485	1.44	1.44	0.31	1.09	2.18	0.38	1.22	1.77	0.26
Hispanic or Latino Black or African American	105	58.10	58.61	23.83	46.67	50.57	18.93	51.43	50.12	15.69
Hispanic or Latino Other	388	9.54	16.94	3.79	6.19	16.53	4.71	8.76	19.19	2.94
Non-Hispanic or Latino White	33,101	0.55	1.01	0.14	0.37	0.86	0.10	0.33	0.72	0.03
Non-Hispanic or Latino Black or African American	5,772	3.10	4.78	0.79	2.20	3.93	0.82	2.36	3.86	0.42
Non-Hispanic or Latino Other	3,042	5.42	8.09	1.78	5.52	9.85	2.28	5.75	9.95	1.41
% Hispanic or Latino in Segment										
50-100%	3,370	2.05	2.91	0.78	2.05	4.22	0.82	2.31	3.87	0.58
10-50%	8,517	1.84	2.57	0.57	1.50	2.15	0.53	1.61	2.23	0.38
<10%	36,006	1.32	2.05	0.38	0.98	1.88	0.36	1.00	1.71	0.18

Table E.2 2005 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	n	Before res.qdu.nr[1] (SDUWT*DU05WT10*DU05WT11) % Unweighted	% Weighted[2]	% Outwinsor[3]	After res.qdu.nr[1] (SDUWT*DU05WT10*...*DU05WT12) % Unweighted	% Weighted[2]	% Outwinsor[3]	Final Weight: After res.qdu.ps[1] (SDUWT*DU05WT10*...*DU05WT13) % Unweighted	% Weighted[2]	% Outwinsor[3]
% Black or African American in Segment										
50-100%	3,820	3.06	4.76	0.78	2.30	4.88	1.15	2.36	4.01	0.56
10-50%	7,414	2.16	3.82	0.90	1.70	2.98	0.68	1.92	3.25	0.43
<10%	36,659	1.16	1.58	0.30	0.92	1.64	0.30	0.94	1.48	0.18
% Owner-Occupied DUs in Segment										
50-100%	35,934	1.23	2.01	0.37	0.92	1.70	0.33	0.92	1.64	0.19
10-50%	9,235	2.11	2.89	0.70	1.59	2.64	0.63	1.72	2.37	0.39
<10%	2,724	2.35	2.69	0.60	2.72	5.84	1.01	3.08	5.15	0.57
Combined Median Rent/Housing Value										
1st Quintile	7,971	1.66	2.25	0.40	0.95	1.40	0.32	1.08	1.45	0.19
2nd Quintile	10,700	1.60	2.49	0.53	0.97	2.34	0.46	1.00	1.98	0.27
3rd Quintile	10,655	1.42	2.18	0.44	1.06	1.75	0.39	1.10	1.82	0.23
4th Quintile	10,233	1.27	2.02	0.44	1.09	1.85	0.44	1.28	1.81	0.24
5th Quintile	8,334	1.42	2.20	0.42	1.74	3.02	0.51	1.61	2.67	0.30
Population Density										
Large MSA[1]	20,074	1.58	2.27	0.44	1.51	2.58	0.56	1.65	2.43	0.33
Medium to Small MSA[1]	23,434	1.37	2.17	0.43	0.96	1.73	0.29	0.94	1.60	0.15
Non-MSA,[1] Urban	1,345	2.01	2.11	0.38	0.67	1.31	0.35	0.82	1.38	0.19
Non-MSA,[1] Rural	3,040	1.22	2.27	0.64	0.43	0.79	0.22	0.36	0.78	0.24
Group Quarters										
Group	822	3.04	4.86	0.65	0.73	1.56	0.31	0.73	1.58	0.31
Nongroup	47,071	1.44	2.20	0.44	1.16	2.12	0.43	1.21	1.98	0.25
Household Size										
One	5,673	1.04	1.89	0.28	0.67	1.69	0.27	0.72	1.15	0.07
Two	17,353	0.99	1.51	0.30	0.86	1.48	0.27	0.90	1.51	0.16
Three	13,581	1.69	3.27	0.71	1.39	3.27	0.80	1.40	3.25	0.51
Four or More	11,286	2.14	4.17	0.97	1.53	4.17	0.98	1.67	4.14	0.69

[1] DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PS = poststratification adjustment, QDU = questionnaire dwelling unit, Res = Respondent, SDU = screener dwelling unit.
[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 nonextreme 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

Table F.1 2005 NSDUH QDU-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	47,893	112,636,350	112,636,350	112,636,350	0.00	-0.00
Census Region						
Northeast	9,617	21,104,671	21,104,671	21,104,671	0.00	-0.00
South	14,744	41,161,319	41,161,319	41,161,319	0.00	-0.00
Midwest	13,342	25,910,507	25,910,507	25,910,507	0.00	-0.00
West	10,190	24,459,853	24,459,853	24,459,853	0.00	-0.00
Quarter						
Quarter 1	11,424	28,362,990	28,362,990	28,362,990	0.00	0.00
Quarter 2	12,800	27,876,402	27,876,402	27,876,402	0.00	-0.00
Quarter 3	11,900	28,118,711	28,118,711	28,118,711	0.00	-0.00
Quarter 4	11,769	28,278,248	28,278,248	28,278,248	0.00	-0.00
Household Type						
12-17, 18-25, 26+	4,634	4,728,244	4,728,244	4,728,244	-0.00	-0.00
12-17, 18-25	111	100,109	100,109	100,109	0.00	0.00
12-17, 26+	14,757	13,993,215	13,993,215	13,993,215	-0.00	-0.00
18-25, 26+	9,477	12,137,100	12,137,100	12,137,100	0.00	0.00
12-17	33	23,585	23,585	23,585	0.00	-0.00
18-25	5,964	6,401,239	6,401,239	6,401,239	0.00	-0.00
26+	12,917	75,252,859	75,252,858	75,252,858	0.00	0.00
Race/Ethnicity of Householder						
Hispanic or Latino White	5,485	11,249,746	11,249,746	11,249,746	0.00	-0.00
Hispanic or Latino Black or African American	105	465,670	460,089	465,670	0.00	-1.20
Hispanic or Latino Other	388	532,965	538,546	532,965	0.00	1.05
Non-Hispanic or Latino White	33,101	81,057,523	81,057,523	81,057,523	0.00	-0.00
Non-Hispanic or Latino Black or African American	5,772	13,312,008	13,312,008	13,312,008	0.00	-0.00
Non-Hispanic or Latino Other	3,042	6,018,438	6,018,438	6,018,438	0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	3,370	8,359,517	8,359,517	8,359,517	0.00	-0.00
10-50%	8,517	22,946,051	22,946,051	22,946,051	0.00	0.00
<10%	36,006	81,330,782	81,330,782	81,330,782	0.00	-0.00
% Black or African American in Segment						
50-100%	3,820	8,332,037	8,332,036	8,332,036	0.00	-0.00
10-50%	7,414	19,764,772	19,764,772	19,764,772	0.00	0.00
<10%	36,659	84,539,542	84,539,542	84,539,542	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	35,934	84,088,418	84,088,418	84,088,418	0.00	-0.00
10-50%	9,235	22,361,671	22,361,671	22,361,671	0.00	-0.00
<10%	2,724	6,186,261	6,186,261	6,186,261	0.00	-0.00

Table F.1 2005 NSDUH QDU-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%
Combined Median Rent/Housing Value						
1st Quintile	7,971	16,481,840	16,481,840	16,481,840	0.00	-0.00
2nd Quintile	10,700	22,950,519	22,950,519	22,950,519	0.00	-0.00
3rd Quintile	10,655	23,913,115	23,913,115	23,913,115	0.00	-0.00
4th Quintile	10,233	25,271,104	25,271,104	25,271,104	0.00	0.00
5th Quintile	8,334	24,019,774	24,019,774	24,019,774	0.00	-0.00
Population Density						
Large MSA	20,074	57,936,414	57,936,414	57,936,414	0.00	-0.00
Medium to Small MSA	23,434	46,493,289	46,493,289	46,493,289	0.00	-0.00
Non-MSA, Urban	1,345	2,386,405	2,386,405	2,386,405	-0.00	0.00
Non-MSA, Rural	3,040	5,820,242	5,820,242	5,820,242	0.00	-0.00
Group Quarters						
Group	822	918,535	918,535	918,535	0.00	0.00
Nongroup	47,071	111,717,815	111,717,815	111,717,815	0.00	-0.00
Household Size						
One	5,673	29,590,028	29,581,004	29,857,979	-0.90	-0.93
Two	17,353	53,029,128	53,030,848	52,777,497	0.48	0.48
Three	13,581	17,549,426	17,562,711	17,390,018	0.92	0.99
Four or More	11,286	12,467,767	12,461,787	12,610,856	-1.13	-1.18

DU = dwelling unit, MSA = metropolitan statistical area, QDU = questionnaire dwelling unit, SDU = screener dwelling unit.

[1] YR05WT1*...*YR05WT9*DU05WT10*...*DU05WT12 (before QDU poststratification).

[2] YR05WT1*...*YR05WT9*DU05WT10*...*DU05WT13 (after QDU poststratification).

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

Table G.1 2005 NSDUH Selected QDU-Level Weight Summary Statistics

		SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)						Before sel.qdu.ps[1] (SDUWT*DU05WT10)						After sel.qdu.ps[1] (SDUWT*DU05WT10*DU05WT11)					
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]**
Total	57,243	17	486	726	1,168	9,584	1.46	17	611	1,043	2,111	35,086	2.76	8	600	1,033	2,139	39,585	2.76
Census Region																			
Northeast	11,599	22	412	669	953	6,971	1.40	22	593	885	1,835	25,305	2.91	8	572	883	1,890	28,144	2.93
South	17,579	36	672	977	1,323	9,584	1.34	36	821	1,294	2,614	31,319	2.51	26	801	1,291	2,608	39,585	2.54
Midwest	15,996	23	519	634	802	6,105	1.31	23	590	790	1,821	35,086	2.62	15	585	802	1,786	29,386	2.59
West	12,069	17	275	677	1,631	6,540	1.66	17	403	1,169	2,205	32,377	2.96	17	393	1,127	2,236	27,246	2.95
Quarter																			
Quarter 1	13,629	21	513	792	1,230	9,584	1.43	21	662	1,122	2,201	31,549	2.71	9	648	1,112	2,214	28,057	2.72
Quarter 2	15,342	17	484	686	1,074	6,483	1.43	17	592	951	1,990	23,231	2.70	15	585	954	2,019	27,027	2.67
Quarter 3	14,157	21	481	699	1,188	6,296	1.49	21	599	1,049	2,121	28,309	2.77	8	582	1,038	2,158	39,585	2.83
Quarter 4	14,115	18	475	730	1,195	9,209	1.48	22	610	1,041	2,137	35,086	2.81	24	604	1,042	2,195	29,386	2.78
Household Type																			
12-17, 18-25, 26+	5,197	22	508	778	1,211	6,709	1.45	22	508	778	1,211	6,709	1.45	9	484	767	1,203	9,362	1.52
12-17, 18-25	132	57	396	662	1,029	2,491	1.44	57	396	662	1,029	2,491	1.44	33	325	654	1,055	3,000	1.57
12-17, 26+	16,751	21	452	694	1,102	9,584	1.49	21	454	699	1,113	9,613	1.49	8	432	695	1,107	9,243	1.54
18-25, 26+	11,310	17	543	761	1,250	6,540	1.44	17	621	874	1,456	7,707	1.44	17	598	877	1,449	8,760	1.48
12-17	35	52	248	575	978	2,347	1.59	57	248	587	978	2,370	1.60	53	261	539	935	2,049	1.61
18-25	6,735	25	441	717	1,144	6,296	1.46	26	509	805	1,326	7,644	1.45	21	493	817	1,272	6,271	1.47
26+	17,083	18	502	719	1,180	6,002	1.44	48	1,931	3,460	5,810	35,086	1.68	25	1,887	3,442	5,836	39,585	1.69
Race/Ethnicity of Householder																			
Hispanic or Latino White	6,467	25	536	872	1,364	5,989	1.36	26	631	1,138	1,943	25,305	2.39	8	615	1,127	1,946	25,193	2.41
Hispanic or Latino Black or African American	117	65	1,051	1,620	2,439	6,971	1.49	86	1,300	2,117	3,564	32,377	2.51	107	1,820	2,630	4,180	33,427	2.30
Hispanic or Latino Other	442	17	142	285	959	6,540	2.78	17	184	394	1,430	14,118	3.58	17	165	441	1,406	17,841	3.55
Non-Hispanic or Latino White	39,900	22	488	696	1,105	6,709	1.42	22	619	1,036	2,188	27,054	2.75	9	610	1,026	2,211	28,584	2.73
Non-Hispanic or Latino Black or African American	6,632	24	594	861	1,244	6,483	1.43	32	703	1,086	2,090	35,086	2.80	23	692	1,077	2,123	29,386	2.85
Non-Hispanic or Latino Other	3,685	22	215	582	1,181	9,584	1.87	22	303	849	1,862	31,319	3.09	15	297	818	1,916	39,585	3.20
% Hispanic or Latino in Segment																			
50-100%	3,983	39	612	1,071	1,543	5,989	1.31	39	764	1,362	2,152	32,377	2.35	28	741	1,408	2,290	25,193	2.33
10-50%	10,169	17	574	928	1,424	6,971	1.37	17	719	1,330	2,395	28,197	2.54	17	709	1,277	2,408	39,585	2.53
<10%	43,091	18	427	681	1,053	9,584	1.48	21	581	946	2,023	35,086	2.85	8	572	945	2,047	33,427	2.86

Table G.1 2005 NSDUH Selected QDU-Level Weight Summary Statistics (continued)

		SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)						Before sel.qdu.ps[1] (SDUWT*DU05WT10)						After sel.qdu.ps[1] (SDUWT*DU05WT10*DU05WT11)					
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 or African American in Segment																			
50-100%	4,430	36	568	813	1,126	6,971	1.43	36	670	999	1,861	35,086	3.00	29	669	1,007	1,976	29,386	2.95
10-50%	8,744	39	622	905	1,374	9,584	1.37	39	746	1,285	2,359	32,377	2.59	27	740	1,269	2,422	39,585	2.64
<10%	44,069	17	426	687	1,122	6,709	1.47	17	581	1,001	2,081	31,319	2.77	8	570	983	2,102	28,584	2.76
% Owner-Occupied DUs[1] in Segment																			
50-100%	43,173	18	474	703	1,128	9,584	1.45	21	604	1,022	2,111	31,549	2.73	8	592	1,013	2,135	39,585	2.75
10-50%	10,916	17	514	797	1,272	9,209	1.47	17	635	1,100	2,133	35,086	2.84	15	631	1,083	2,170	29,386	2.79
<10%	3,154	25	535	830	1,307	5,974	1.45	26	654	1,096	2,052	25,208	2.75	21	650	1,115	2,130	25,193	2.71
Combined Median Rent/Housing Value																			
1st Quintile	9,237	21	393	657	1,021	9,209	1.50	21	551	910	1,919	31,549	2.78	8	545	916	2,020	28,057	2.79
2nd Quintile	12,537	18	422	682	1,056	6,971	1.49	25	570	935	1,976	32,377	2.92	12	560	928	1,984	33,427	2.89
3rd Quintile	12,717	17	437	704	1,151	6,709	1.48	17	593	1,006	2,062	27,816	2.84	9	564	974	2,029	39,585	2.86
4th Quintile	12,406	23	511	756	1,233	9,584	1.44	28	638	1,120	2,178	35,086	2.70	24	626	1,085	2,162	29,386	2.69
5th Quintile	10,346	23	586	863	1,305	5,971	1.37	23	710	1,266	2,473	28,197	2.50	18	710	1,295	2,520	27,246	2.52
Population Density																			
Large MSA[1]	24,515	36	646	913	1,384	6,971	1.32	36	767	1,323	2,515	35,086	2.50	26	760	1,313	2,561	39,585	2.49
Medium to Small MSA[1]	27,647	17	339	619	977	9,584	1.54	17	485	846	1,819	26,698	2.92	8	473	848	1,823	28,584	2.96
Non-MSA,[1] Urban	1,536	28	286	576	911	3,049	1.53	29	403	821	1,654	24,081	3.01	17	414	806	1,681	20,649	2.95
Non-MSA,[1] Rural	3,545	22	262	582	917	6,002	1.58	22	398	807	1,720	31,319	3.04	9	401	802	1,751	33,427	3.04
Group Quarters																			
Group	854	25	316	701	1,014	4,137	1.59	26	351	783	1,177	26,698	3.07	21	336	790	1,181	28,007	3.09
Nongroup	56,389	17	487	727	1,171	9,584	1.46	17	613	1,050	2,127	35,086	2.75	8	602	1,041	2,158	39,585	2.75
Household Size																			
One	7,114	26	438	686	1,106	6,296	1.44	26	889	2,410	6,092	35,086	2.19	21	894	2,401	6,135	29,386	2.17
Two	21,333	17	490	717	1,156	9,209	1.45	17	721	1,530	3,310	28,309	2.15	17	720	1,502	3,308	39,585	2.17
Three	15,900	22	496	732	1,164	9,584	1.45	22	534	811	1,373	13,929	1.94	12	520	805	1,365	16,476	1.95
Four or More	12,896	21	486	755	1,229	6,971	1.49	21	496	776	1,277	9,822	1.62	8	475	767	1,268	13,761	1.68

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

Table G.2 2005 NSDUH Respondent QDU-Level Weight Summary Statistics

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DU05WT10*DU05WT11)						After res.qdu.nr[1] (SDUWT*DU05WT10*...*DU05WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU05WT10*...*DU05WT13)					
		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	47,893	8	582	985	1,982	39,585	2.78	9	659	1,129	2,380	43,712	3.16	8	660	1,132	2,381	44,318	3.15
Census Region																			
Northeast	9,617	8	550	850	1,698	26,685	2.99	9	612	979	2,104	41,341	3.45	8	611	979	2,104	37,820	3.46
South	14,744	26	790	1,252	2,299	39,585	2.56	27	887	1,429	2,749	43,695	2.89	24	887	1,430	2,753	44,318	2.89
Midwest	13,342	15	571	774	1,653	28,057	2.61	15	648	914	2,028	43,712	2.92	17	649	915	2,030	35,881	2.91
West	10,190	17	376	1,067	2,150	26,488	2.94	20	428	1,237	2,510	43,698	3.36	20	429	1,237	2,510	42,835	3.37
Quarter																			
Quarter 1	11,424	9	622	1,066	2,042	28,057	2.77	13	702	1,215	2,460	43,712	3.14	12	704	1,217	2,462	40,084	3.13
Quarter 2	12,800	15	569	907	1,844	27,027	2.70	15	643	1,051	2,258	41,739	3.05	17	643	1,051	2,265	43,205	3.05
Quarter 3	11,900	8	567	992	2,012	39,585	2.82	9	640	1,122	2,376	43,698	3.26	8	642	1,126	2,377	42,835	3.25
Quarter 4	11,769	24	585	996	2,032	28,007	2.80	25	673	1,142	2,443	43,695	3.14	23	674	1,144	2,435	44,318	3.15
Household Type																			
12-17, 18-25, 26+	4,634	9	476	759	1,195	9,362	1.53	13	524	850	1,352	11,016	1.55	12	524	850	1,351	13,605	1.55
12-17, 18-25	111	33	320	644	1,044	3,000	1.61	33	362	731	1,234	3,073	1.57	32	365	744	1,231	3,337	1.63
12-17, 26+	14,757	8	424	693	1,105	9,243	1.55	9	468	785	1,250	12,470	1.58	8	469	788	1,255	10,960	1.56
18-25, 26+	9,477	17	596	880	1,456	8,760	1.49	20	688	1,041	1,719	9,539	1.52	20	689	1,046	1,724	7,800	1.50
12-17	33	53	263	539	935	2,049	1.61	56	263	602	938	2,091	1.61	55	228	601	925	2,096	1.63
18-25	5,964	21	486	809	1,255	6,271	1.47	21	547	909	1,449	6,983	1.51	22	548	908	1,451	6,778	1.51
26+	12,917	25	1,838	3,366	5,675	39,585	1.69	25	2,337	4,421	7,604	43,712	1.76	23	2,329	4,416	7,613	44,318	1.76
Race/Ethnicity of Householder																			
Hispanic or Latino White	5,485	8	602	1,092	1,851	25,193	2.40	9	683	1,247	2,186	43,698	2.79	8	683	1,253	2,189	42,835	2.79
Hispanic or Latino Black or African American	105	107	1,810	2,540	4,014	33,427	2.38	107	1,858	2,841	4,590	29,457	2.24	95	1,840	2,718	4,366	31,739	2.41
Hispanic or Latino Other	388	17	162	428	1,227	10,798	3.20	20	190	475	1,490	19,932	3.60	20	185	483	1,482	20,103	3.56
Non-Hispanic or Latino White	33,101	9	592	972	2,029	28,007	2.77	13	674	1,123	2,467	43,695	3.12	12	675	1,125	2,471	43,964	3.12
Non-Hispanic or Latino Black or African American	5,772	23	679	1,059	2,012	28,057	2.83	23	743	1,167	2,276	43,712	3.22	25	746	1,170	2,279	44,318	3.23
Non-Hispanic or Latino Other	3,042	15	265	755	1,777	39,585	3.33	15	310	886	2,153	42,891	3.89	17	312	885	2,151	42,815	3.89
% Hispanic or Latino in Segment																			
50-100%	3,370	28	720	1,357	2,189	25,193	2.35	28	815	1,561	2,574	43,698	2.66	27	815	1,566	2,590	42,835	2.66
10-50%	8,517	17	690	1,236	2,268	39,585	2.57	20	789	1,413	2,696	42,891	2.90	20	789	1,417	2,705	42,815	2.91
<10%	36,006	8	553	904	1,875	33,427	2.88	9	627	1,042	2,271	43,712	3.28	8	627	1,043	2,273	44,318	3.27

Table G.2 2005 NSDUH Respondent QDU-Level Weight Summary Statistics (continued)

		Before res.qdu.nr[1] (SDUWT*DU05WT10*DU05WT11)						After res.qdu.nr[1] (SDUWT*DU05WT10*...*DU05WT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DU05WT10*...*DU05WT13)					
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 or African American in Segment																			
50-100%	3,820	32	663	991	1,864	28,057	2.81	32	733	1,114	2,165	43,712	3.25	27	732	1,123	2,155	43,205	3.23
10-50%	7,414	27	722	1,219	2,248	39,585	2.68	28	814	1,409	2,690	42,891	2.95	24	816	1,409	2,701	44,318	2.96
<10%	36,659	8	550	933	1,933	28,007	2.79	9	625	1,076	2,339	43,698	3.19	8	625	1,080	2,339	43,964	3.19
% Owner-Occupied DUs[1] in Segment																			
50-100%	35,934	8	572	962	1,961	39,585	2.80	9	651	1,104	2,377	43,712	3.17	8	651	1,106	2,379	44,318	3.17
10-50%	9,235	15	613	1,039	2,048	27,781	2.76	15	684	1,183	2,407	40,864	3.13	17	682	1,187	2,407	41,039	3.14
<10%	2,724	21	644	1,090	2,016	25,193	2.66	21	713	1,242	2,338	43,698	3.06	22	715	1,251	2,356	42,835	3.06
Combined Median Rent/Housing Value																			
1st Quintile	7,971	8	528	885	1,884	28,057	2.74	9	591	998	2,190	43,712	3.06	8	590	999	2,190	35,881	3.05
2nd Quintile	10,700	12	547	899	1,843	33,427	2.94	12	610	1,020	2,163	43,698	3.31	12	609	1,018	2,162	43,964	3.31
3rd Quintile	10,655	9	548	944	1,905	39,585	2.86	13	620	1,081	2,267	42,891	3.22	12	621	1,081	2,274	42,815	3.22
4th Quintile	10,233	24	604	1,032	2,002	27,781	2.75	26	693	1,195	2,432	42,874	3.12	24	694	1,201	2,430	44,318	3.12
5th Quintile	8,334	18	688	1,228	2,313	26,488	2.55	25	811	1,449	2,923	41,740	2.91	33	812	1,456	2,918	42,020	2.91
Population Density																			
Large MSA[1]	20,074	26	746	1,256	2,357	39,585	2.51	27	864	1,458	2,920	43,712	2.85	27	867	1,463	2,911	42,835	2.85
Medium to Small MSA[1]	23,434	8	455	819	1,713	28,007	2.96	9	513	940	2,019	43,695	3.35	8	513	940	2,021	44,318	3.34
Non-MSA,[1] Urban	1,345	46	405	785	1,602	20,649	2.98	47	444	870	1,781	28,263	3.32	46	442	871	1,784	28,191	3.32
Non-MSA,[1] Rural	3,040	9	378	766	1,618	33,427	3.11	13	424	889	1,936	29,457	3.33	12	424	886	1,936	31,739	3.35
Group Quarters																			
Group	822	21	336	784	1,152	28,007	3.02	21	358	817	1,211	43,695	4.13	22	360	810	1,209	43,964	4.16
Nongroup	47,071	8	584	992	2,003	39,585	2.77	9	663	1,140	2,410	43,712	3.14	8	664	1,144	2,410	44,318	3.14
Household Size																			
One	5,673	21	844	2,124	5,632	28,057	2.24	21	983	2,600	7,536	43,712	2.39	22	978	2,617	7,519	43,964	2.38
Two	17,353	17	698	1,390	3,032	39,585	2.22	20	806	1,645	3,888	42,891	2.43	20	806	1,647	3,878	44,318	2.44
Three	13,581	12	509	794	1,343	16,476	1.91	12	571	924	1,555	22,877	2.15	12	572	926	1,560	22,997	2.15
Four or More	11,286	8	468	760	1,249	13,761	1.66	9	517	857	1,420	27,951	1.85	8	515	858	1,424	29,342	1.86

[1] DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PS = poststratification adjustment, QDU = questionnaire dwelling unit, Res = respondent, SDU = screener dwelling unit, Sel = selected.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is 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 2005, 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 H.1a through H.2b. 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 led to situations whereby modeling inclusion of all originally proposed levels of covariates in the model was not possible. The text and exhibits in Sections H.1 and H.2 summarize the decisions made with regard to final covariates included in each model. For the list of proposed initial covariates considered at each stage of modeling, see Exhibit H.2. For the list of realized final model covariates, see Exhibits H.1.1 to H.2.4. For guidelines on interpreting these exhibits, see Appendix C.

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). A complete list of all variables and associated levels used at any stage of modeling is provided. 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,[1] 3: DU with 3 persons, 4: DU with ≥ 4 persons

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

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

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

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

Pair Race/Ethnicity (10 Levels)
1: white and white,[1] 2: white and black or African American, 3: white and Hispanic or Latino, 4: white and other, 5: black or African American and black or African American, 6: black or African American and Hispanic or Latino, 7: black or African American and other, 8: Hispanic or Latino and Hispanic or Latino, 9: Hispanic or Latino and other, 10: other and other

Pair Race/Ethnicity (5 Levels)
1: Mixed race pair, 2: Hispanic or Latino pair, 3: black or African American pair, 4: white pair,[1] 5: other pair

Pair Race/Ethnicity (4 Levels)
1: Mixed race pair or other and other, 2: Hispanic or Latino pair, 3: black or African American 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 or African American (% Black)
1: 50-100%, 2: 10->50%, 3: 0->10%[1]

Percentage of Segments That Are Hispanic or Latino (% 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/Ethnicity of Householder
1: Hispanic or Latino white,[1] 2: Hispanic or Latino black or African American, 3: Hispanic or Latino others, 4: Non-Hispanic or Latino white, 5: Non-Hispanic or Latino black or African American, 6: Non-Hispanic or Latino 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;[1] 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming; 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/partner-partner
10: Spouse-spouse/partner-partner with children (younger than 18)

DU = dwelling unit, MSA = metropolitan statistical area.

[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.

* The pair member focused on.

Exhibit H.2 Covariates for 2005 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 Gender	4	2
Pair Race/Ethnicity	10	9
Race/Ethnicity of Householder	6	5
Rent/Housing	5	4
Segment % Black or African American	3	2
Segment % Hispanic or Latino	3	2
% Owner-Occupied	3	2
Pair Relationship	Model-specific	
Two-Factor Effects		
Pair Race/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8
Pair Gender x Pair Age (6 Levels)	3 x 6	10
State/Region x Pair Race/Ethnicity (5 Levels)	Model-specific	
State/Region x Pair Age (6 Levels)	Model-specific	
State/Region x Pair Gender	Model-specific	
Rent/Housing x % Black or African American	5 x 3	8
Rent/Housing x % Hispanic or Latino	5 x 3	8
Rent/Housing x % Owner-Occupied	5 x 3	8
% Owner-Occupied x % Black or African American	3 x 3	4
% Owner-Occupied x % Hispanic or Latino	3 x 3	4
Three-Factor Effects		
Pair Race/Ethnicity (4 Levels) x Pair Gender x Pair Age (3 Levels)	4 x 3 x 3	12

Appendix H.1: 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 H.1a 2005 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Winsorized			Nominal	Realized
sel.pr.ps	5.86	15.15	5.25	8.043	213	(0.26, 2.05)	(0.26, 2.05)
	2.20	7.42	1.36	7.394	205	(0.20, 3.76)	(0.20, 3.76)
						(0.50, 1.45)	(0.50, 1.44)
res.pr.nr	1.96	7.26	1.63	8.644	213	(1.00, 3.00)	(1.00, 3.00)
	1.93	8.99	1.94	9.606	212	(1.00, 4.93)	(1.00, 4.91)
						(1.14, 1.63)	(1.14, 1.62)
res.pr.ps	1.98	9.54	2.62	9.606	223	(0.35, 2.21)	(0.36, 2.18)
	1.80	8.34	2.17	10.036	215	(0.20, 2.21)	(0.20, 2.17)
						N/A	N/A
res.pr.ev	1.80	8.34	2.17	10.036	223	(0.80 , 2.12)	(0.93, 2.00)
	0.99	5.20	0.95	9.148	215	(0.67, 2.12)	(0.77, 1.64)
						N/A	N/A

GEM = generalized exponential model.

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

[2] Unequal weighting effect (UWE) 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 H.1b 2005 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Midwest and West)

	SDU Weight	Pair Selection Prob		sel.pr.ps[1]		res.pr.nr[1]		res.pr.ps[1]		res.pr.ev[1]	
	1-10	**pairwt11**	**1-11**	**pairwt12**	**1-12**	**pairwt13**	**1-13**	**pairwt14**	**1-14**	**pairwt15**	**1-15**
Minimum	22	1.02	45	0.05	11	0.46	11	0.11	3	0.54	3
1%	72	1.02	160	0.23	96	0.99	102	0.28	62	0.83	56
5%	146	1.18	399	0.32	241	1.01	243	0.50	204	0.88	193
10%	243	1.25	697	0.41	438	1.02	449	0.61	363	0.90	342
25%	575	1.45	1,251	0.64	1,065	1.07	1,118	0.85	1,024	0.95	992
Median	854	4.80	3,148	0.98	3,034	1.16	3,326	1.00	3,244	0.99	3,196
75%	1,223	10.15	8,166	1.32	8,125	1.31	9,251	1.12	9,065	1.02	9,014
90%	1,625	17.48	17,287	1.69	19,516	1.60	24,687	1.34	24,667	1.06	25,347
95%	1,961	27.04	25,957	1.98	32,042	1.89	41,933	1.47	43,283	1.08	44,352
99%	2,774	49.95	57,325	2.63	72,232	2.91	112,716	1.71	112,935	1.17	113,332
Maximum	9,209	4,750.92	1,320,638	3.76	799,764	4.91	1,333,226	2.17	1,420,149	1.64	1,088,810
n	13,227	-	13,227	-	13,227	-	10,168	-	10,168	-	10,168
Mean	932	8.55	7,466	1.03	8,233	1.27	10,710	0.99	10,710	0.99	10,710
Max/Mean	10	-	177	-	97	-	124	-	133	-	102

SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

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/ethnicity by pair gender by pair age was simplified by collapsing. Here, pair race/ethnicity categories were mixed, and other pair, Hispanic or Latino pair, and black or African-American 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

In the respondent pair-level nonresponse step, 212 of 213 proposed factors were retained in the final model.

All main effects and two-factor effects were retained, but three-factor interaction of pair race/ethnicity by pair gender by pair age was collapsed by combining the pair age groups of 12 to 17 and 18+ for "mixed race/ethnicity pair or other and other race/ethnicity" among female-female pairs, resulting in 11 of 12 proposed three-factor effects.

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/ethnicity by pair gender by pair age was simplified by collapsing. Here, pair race/ethnicity categories were mixed, and other pair, Hispanic or Latino pair, and black or African American 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 Extreme Value Adjustment

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

Exhibit H.1.1 Covariates for 2005 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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**4**	
Pair Race/Ethnicity (4 Levels) x Pair Gender 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 gender and pair age.
Total		**213**	**205**	

Exhibit H.1.2 Covariates for 2005 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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) x Pair Gender x Pair Age (3 Levels)	4 x 3 x 3	12	11	Coll. (1,2,2) & (1,2,3); conv.
Total		**213**	**212**	

Exhibit H.1.3 Covariates for 2005 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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	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/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**8**	
Pair Race/Ethnicity (4 Levels) x Pair Gender 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 gender and pair age.
Total		**223**	**215**	

Exhibit H.1.4 Covariates for 2005 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 H.1.3.

Appendix H.2: 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 H.2a 2005 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Winsorized			Nominal	Realized
sel.pr.ps	6.38	19.84	7.66	36.092	212	(0.20, 2.20)	(0.20, 2.12)
	2.40	8.36	1.06	13.039	198	(0.20, 3.05)	(0.20, 2.94)
						(0.42, 1.88)	(0.43, 1.83)
res.pr.nr	2.35	10.63	1.26	16.585	212	(1.00, 2.90)	(1.00, 2.80)
	2.43	8.61	1.32	13.775	211	(1.00, 4.78)	(1.00, 4.63)
						(1.00, 5.00)	N/A
res.pr.ps	2.55	15.23	5.60	13.775	222	(0.20, 1.80)	(0.20, 1.76)
	2.07	10.76	1.62	9.417	210	(0.20, 2.88)	(0.21, 2.79)
						(0.95, 1.10)	(0.95, 0.95)
res.pr.ev	2.07	10.76	1.62	9.417	222	(0.97, 1.38)	(0.98, 1.37)
	1.01	7.82	0.83	9.290	210	(0.85, 2.00)	(0.90, 1.20)
						(0.96, 1.50)	(0.98, 0.98)

GEM = generalized exponential model.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.
[2] Unequal weighting effect (UWE) 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 H.2b 2005 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

	SDU Weight	Pair Selection		sel.pr.ps[1]		res.pr.nr[1]		res.pr.ps[1]		res.pr.ev[1]	
	1-10	**pairwt11**	**1-11**	**pairwt12**	**1-12**	**pairwt13**	**1-13**	**pairwt14**	**1-14**	**pairwt15**	**1-15**
Minimum	17	1.02	33	0.07	15	0.52	18	0.11	19	0.66	19
1%	80	1.02	145	0.27	127	0.97	137	0.29	102	0.88	100
5%	129	1.18	298	0.45	314	1.02	340	0.56	272	0.95	269
10%	181	1.23	548	0.57	523	1.04	578	0.65	500	0.96	492
25%	428	1.48	945	0.81	1,025	1.08	1,154	0.82	1,046	0.98	1,044
Median	649	4.45	2,611	1.06	2,580	1.18	2,952	0.98	2,881	1.00	2,882
75%	1,128	9.95	7,108	1.33	6,857	1.36	7,930	1.15	7,919	1.02	7,932
90%	1,812	17.49	15,522	1.63	15,706	1.66	19,988	1.34	20,747	1.04	20,772
95%	2,097	25.98	23,004	1.87	25,586	1.95	34,139	1.48	34,898	1.06	34,730
99%	2,599	47.18	57,575	2.32	60,525	2.80	88,597	1.93	99,291	1.10	96,908
Maximum	6,540	1,789.80	2,655,517	11.32	1,195,500	4.63	1,330,080	2.79	860,030	1.30	825,527
n	13,335	-	13,335	-	13,335	-	10,247	-	10,247	-	10,247
Mean	835	8.33	7,481	1.09	7,103	1.29	9,244	1.00	9,244	1.00	9,244
Max/Mean	8	-	355	-	168	-	144	-	93	-	89

GEM = generalized exponential model, SDU = screener dwelling unit.
[1] For a key to modeling abbreviations, see Chapter 7, Exhibit 7.1.

Model Group 2 Overview

Selected Pair-Level Poststratification

In the selected pair-level poststratification step, 73 of 75 proposed main effects were included in the model. The race pairs of "black or African American and Hispanic or Latino" and "black or African American and other" were combined. The race of householder categories of "Hispanic or Latino black or African American" and "Hispanic or Latino others" also were combined. Both sets of variables were collapsed to obtain convergence at acceptable bounds.

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

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

Respondent Pair-Level Nonresponse

As in the previous step, the main-effect race pair categories of "black or African American and Hispanic or Latino" and "black or African American and other" were combined, leaving 74 of 75 initially proposed main effects.

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

All 12 proposed three-factor effects were kept in the model.

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 exactly the same variables as in the respondent pair-level poststratification step.

Exhibit H.2.1 Covariates for 2005 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**73**	
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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	8	Coll. (6) & (7); conv.
Race/Ethnicity of Householder	6	5	4	Coll. (2) & (3); conv
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) x Pair Gender x Pair Age (3 Levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**212**	**198**	

Exhibit H.2.2 Covariates for 2005 NSDUH Pair Weights (res.pr.nr) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**74**	
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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	8	Coll. (6) & (7); conv.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) x Pair Gender x Pair Age (3 Levels)	4 x 3 x 3	12	12	All levels present.
Total		**212**	**211**	

Exhibit H.2.3 Covariates for 2005 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 Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	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/Ethnicity (5 Levels) x Pair Age (6 Levels)	5 x 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) x Pair Gender	5 x 3	8	8	All levels present.
Pair Gender x Pair Age (6 Levels)	3 x 6	10	10	All levels present.
State/Region x Pair Race/Ethnicity (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 Gender	6 x 3	10	10	All levels present.
Rent/Housing x % Black or African American	5 x 3	8	8	All levels present.
Rent/Housing x % Hispanic or Latino	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 or African American	3 x 3	4	4	All levels present.
% Owner-Occupied x % Hispanic or Latino	3 x 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) x Pair Gender x Pair Age (3 Levels)	4 x 3 x 3	12	0	Drop all; conv.
Total		**222**	**210**	

Exhibit H.2.4 Covariates for 2005 NSDUH Pair Weights (res.pr.ev) Model Group 2: Midwest and West

This step used the same variables as the respondent pair-level poststratification step in Exhibit H.2.3.

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

Table I.1 2005 NSDUH Person Pair-Level Response Rates

Domain	Selected Pair Size	Respondent Pair Size	% Interview Response Rate[1]
Total	26,562	20,415	70.37
Pair Age Group			
12-17, 12-17	4,476	3,915	86.50
12-17, 18-25	3,283	2,658	81.57
12-17, 26-34	879	719	82.96
12-17, 35-49	4,187	3,293	78.93
12-17, 50+	623	450	70.41
18-25, 18-25	5,508	4,334	78.65
18-25, 26-34	1,206	847	67.56
18-25, 35-49	1,609	1,125	70.12
18-25, 50+	763	504	64.34
26-34, 26-34	957	667	67.70
26-34, 35-49	563	373	68.98
26-34, 50+	220	133	63.63
35-49, 35-49	867	565	71.93
35-49, 50+	447	253	61.53
50+, 50+	974	579	58.78
Pair Race/Ethnicity			
Hispanic or Latino	3,587	2,738	73.18
Black or African American	2,706	2,239	76.74
White	16,779	12,863	69.81
Other	1,527	1,085	55.94
White & Black or African American	195	157	83.95
White & Hispanic or Latino	835	625	74.15
White & Other	640	478	68.81
Black or African American & Hispanic or Latino	82	68	54.75
Black or African American & Other	81	60	76.00
Hispanic or Latino & Other	130	102	71.15
Pair Gender			
Male, Male	5,769	4,346	67.21
Female, Female	5,629	4,618	75.61
Male, Female	15,164	11,451	69.85
Household Size			
Two	7,569	5,555	64.46
Three	7,194	5,509	71.31
Four or More	11,799	9,351	73.04

Table I.1 2005 NSDUH Person Pair-Level Response Rates (continued)

Domain	Selected Pairs	Respondent Pairs	% Interview Response Rate[1]
Census Region			
Northeast	5,395	4,094	68.24
South	7,832	6,074	70.35
Midwest	7,546	5,812	70.28
West	5,789	4,435	72.07
Quarter			
Quarter 1	6,375	4,956	72.00
Quarter 2	7,100	5,369	66.96
Quarter 3	6,569	5,063	72.26
Quarter 4	6,518	5,027	70.33
% Hispanic or Latino in Segment			
50-100%	2,156	1,641	76.32
10-50%	4,813	3,633	66.87
<10%	19,593	15,141	70.45
% Black or African American in Segment			
50-100%	1,955	1,557	71.48
10-50%	4,038	3,156	73.31
<10%	20,569	15,702	69.60
% Owner-Occupied DUs in Segment			
50-100%	20,312	15,570	69.96
10-50%	4,885	3,780	72.01
<10%	1,365	1,065	73.01
Combined Median Rent/Housing Value			
1st Quintile	4,295	3,456	74.85
2nd Quintile	5,860	4,646	74.28
3rd Quintile	5,828	4,480	70.48
4th Quintile	5,742	4,283	68.33
5th Quintile	4,837	3,550	65.72
Population Density			
Large MSA	11,496	8,507	68.30
Medium to Small MSA	12,771	10,076	72.40
Non-MSA, Urban	659	536	78.69
Non-MSA, Rural	1,636	1,296	74.95
Group Quarters			
Group	472	429	90.58
Nongroup	26,090	19,986	70.27

DU = dwelling unit, MSA = metropolitan statistical area.

[1] The weight used for calculating the response rate includes screener dwelling unit (SDU)- and pair-level design weights, SDU nonresponse and poststratification adjustments, and selected pair poststratification adjustment. This weight is the product of YR05WT1*..*YR05WT9*PR05WT10*PR05WT11.

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

Table J.1 2005 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors

		SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)			Before sel.pr.ps[1] (SDUWT*PR05WT10)			After sel.pr.ps[1] (SDUWT*PR05WT10*PR05WT11)		
Domain	***n***	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**
Total	26,562	1.75	3.35	0.74	6.30	25.11	14.51	2.48	15.50	7.35
Pair Age Group										
12-17, 12-17	4,476	1.16	2.30	0.62	6.48	22.37	7.71	1.43	6.33	0.85
12-17, 18-25	3,283	1.95	3.52	0.68	10.94	33.00	14.30	1.95	5.88	0.91
12-17, 26-34	879	2.05	2.94	0.91	1.71	4.28	1.18	2.39	7.45	1.82
12-17, 35-49	4,187	1.77	3.18	0.64	2.67	8.83	2.70	0.93	3.01	0.52
12-17, 50+	623	1.44	2.70	0.43	1.93	8.49	3.32	1.12	4.58	0.35
18-25, 18-25	5,508	1.83	4.11	0.96	9.10	31.28	13.80	3.27	12.28	1.67
18-25, 26-34	1,206	2.65	5.29	1.35	2.49	13.43	5.30	1.82	7.04	1.22
18-25, 35-49	1,609	2.05	3.57	0.84	5.97	18.39	4.93	3.05	9.69	1.57
18-25, 50+	763	1.57	2.95	0.66	2.75	12.36	5.16	1.44	4.32	0.68
26-34, 26-34	957	2.40	2.94	0.55	4.18	13.03	2.51	1.78	12.62	3.76
26-34, 35-49	563	3.37	4.92	0.86	5.33	49.20	38.28	7.82	32.71	13.29
26-34, 50+	220	1.82	4.79	0.35	3.64	28.27	22.76	3.64	11.07	5.25
35-49, 35-49	867	1.04	2.47	0.56	7.38	56.69	41.66	7.84	53.94	35.23
35-49, 50+	447	1.79	3.83	0.65	5.37	34.98	24.55	3.13	26.57	18.29
50+, 50+	974	0.72	1.31	0.12	7.29	30.25	18.77	5.13	15.74	6.68
Pair Race/Ethnicity										
Hispanic or Latino	3,587	3.40	5.51	1.60	6.41	27.40	14.79	3.32	19.51	9.02
Black or African American	2,706	2.81	5.78	1.13	8.43	24.62	11.47	4.58	16.65	5.36
White	16,779	0.46	1.02	0.16	5.55	21.10	12.27	1.44	12.98	6.84
Other	1,527	5.70	10.24	2.19	8.78	28.59	13.55	6.42	20.27	9.06
White & Black or African American	195	4.62	8.42	1.89	10.26	45.08	31.33	3.08	20.74	5.42
White & Hispanic or Latino	835	3.59	5.20	1.19	6.35	32.08	17.49	2.51	16.81	8.90
White & Other	640	6.25	12.03	2.65	7.81	27.79	17.17	4.69	23.50	9.79
Black or African American & Hispanic or Latino	82	9.76	20.37	6.00	14.63	90.66	84.24	13.41	65.41	31.94
Black or African American & Other	81	3.70	4.05	1.96	4.94	19.68	8.95	1.23	1.16	0.06
Hispanic or Latino & Other	130	9.23	21.37	2.24	8.46	59.40	24.56	5.38	16.71	1.09
Pair Gender										
Male, Male	5,769	2.63	4.95	0.92	8.63	27.41	13.94	3.24	13.44	4.39
Female, Female	5,629	1.39	2.72	0.54	6.54	16.30	5.67	2.38	9.39	1.43
Male, Female	15,164	1.55	2.94	0.74	5.32	26.72	16.95	2.22	17.72	9.77
Household Size										
Two	7,569	1.55	3.32	0.81	0.92	2.12	0.37	0.36	1.10	0.28
Three	7,194	1.46	2.50	0.51	3.11	29.23	20.83	2.45	19.76	9.44
Four or More	11,799	2.06	3.85	0.83	11.69	35.98	19.12	3.86	20.95	10.02

Table J.1 2005 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9) % Unweighted	% Weighted[2]	% Outwinsor[3]	Before sel.pr.ps[1] (SDUWT*PR05WT10) % Unweighted	% Weighted[2]	% Outwinsor[3]	After sel.pr.ps[1] (SDUWT*PR05WT10*PR05WT11) % Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	5,395	1.56	4.02	0.89	6.56	25.10	13.05	2.58	18.08	8.06
South	7,832	1.57	3.26	0.72	5.66	17.77	7.22	2.22	12.28	4.54
Midwest	7,546	1.97	3.43	0.82	6.90	27.93	17.34	2.78	16.80	7.75
West	5,789	1.88	2.91	0.59	6.13	31.23	21.53	2.33	16.88	10.31
Quarter										
Quarter 1	6,375	1.80	3.03	0.63	6.26	24.37	13.67	2.40	19.38	10.65
Quarter 2	7,100	1.38	2.32	0.41	5.75	20.87	10.28	2.63	13.62	5.40
Quarter 3	6,569	1.86	3.98	0.96	6.12	31.48	21.53	2.25	16.01	8.13
Quarter 4	6,518	1.99	4.05	0.96	7.12	22.90	11.64	2.61	13.10	5.31
% Hispanic or Latino in Segment										
50-100%	2,156	1.76	2.84	0.60	6.40	34.35	23.04	3.43	27.08	15.92
10-50%	4,813	2.41	4.13	1.02	6.32	22.33	10.38	2.72	13.74	5.16
<10%	19,593	1.59	3.18	0.67	6.28	24.27	14.21	2.31	14.08	6.57
% Black or African American in Segment										
50-100%	1,955	2.81	6.34	1.45	7.57	19.29	7.09	5.68	17.73	3.90
10-50%	4,038	1.96	3.59	0.90	7.03	24.01	10.50	2.48	15.82	6.76
<10%	20,569	1.61	2.97	0.63	6.03	25.85	15.98	2.17	15.20	7.83
% Owner-Occupied DUs[1] in Segment										
50-100%	20,312	1.57	2.86	0.61	6.02	24.31	14.00	2.31	15.35	7.41
10-50%	4,885	2.37	4.62	1.14	7.25	28.08	16.92	3.64	17.08	7.60
<10%	1,365	2.27	5.19	1.04	7.03	27.49	13.62	0.81	6.58	1.81
Combined Median Rent/Housing Value										
1st Quintile	4,295	1.61	3.00	0.75	6.73	21.55	10.49	2.70	12.31	4.89
2nd Quintile	5,860	1.69	3.01	0.65	6.35	24.14	12.72	2.51	15.47	6.16
3rd Quintile	5,828	1.82	3.32	0.71	6.19	29.11	18.92	1.97	18.23	9.60
4th Quintile	5,742	1.81	3.39	0.76	6.22	27.40	17.71	2.84	16.63	8.53
5th Quintile	4,837	1.80	3.92	0.82	6.08	21.29	10.29	2.42	13.81	6.67
Population Density										
Large MSA[1]	11,496	1.91	3.63	0.85	6.58	28.77	17.24	2.95	18.27	9.73
Medium to Small MSA[1]	12,771	1.73	3.32	0.66	6.22	21.70	12.20	1.99	12.25	4.81
Non-MSA,[1] Urban	659	1.06	1.69	0.58	4.55	10.13	2.69	2.43	8.84	0.99
Non-MSA,[1] Rural	1,636	1.10	1.11	0.21	5.68	14.13	4.61	3.00	12.08	2.52
Group Quarters										
Group	472	3.39	10.22	1.35	10.81	34.47	17.72	4.45	18.17	1.85
Nongroup	26,090	1.72	3.23	0.73	6.22	25.07	14.50	2.44	15.49	7.37

[1] This step used demographic variables from screener data for all selected person pairs; DU = dwelling unit, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, SDU = screener dwelling unit, Sel = selected.
[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 nonextreme 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   2005 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	n	Before res.pr.nr[1] (SDUWT*PR05WT10*PR05WT11) % Unweighted	% Weighted[2]	% Outwinsor[3]	After res.pr.nr[1] (SDUWT*PR05WT10*...*PR05WT12) % Unweighted	% Weighted[2]	% Outwinsor[3]
Total	20,415	2.34	17.67	8.80	2.34	16.92	7.50
Pair Age Group							
12-17, 12-17	3,915	1.23	5.47	0.64	0.51	2.59	0.50
12-17, 18-25	2,658	1.77	5.65	0.95	1.66	6.26	1.49
12-17, 26-34	719	2.09	7.16	1.97	1.25	4.53	1.23
12-17, 35-49	3,293	0.97	3.62	0.60	0.76	2.45	0.67
12-17, 50+	450	1.11	3.81	0.40	0.22	0.68	0.01
18-25, 18-25	4,334	2.91	12.09	1.73	3.41	14.10	3.10
18-25, 26-34	847	2.13	7.35	1.29	4.25	13.65	2.61
18-25, 35-49	1,125	3.73	11.53	1.66	4.18	10.40	1.67
18-25, 50+	504	1.98	3.74	0.87	3.17	10.32	2.13
26-34, 26-34	667	1.35	13.51	4.01	4.35	24.73	5.64
26-34, 35-49	373	7.51	38.71	16.72	6.70	33.53	14.07
26-34, 50+	133	6.02	22.16	8.81	3.76	15.45	4.71
35-49, 35-49	565	8.14	59.49	39.03	7.08	56.45	32.87
35-49, 50+	253	3.95	36.46	27.69	3.95	34.62	21.19
50+, 50+	579	5.70	19.94	9.88	3.80	12.14	5.58
Pair Race/Ethnicity							
Hispanic or Latino	2,738	3.29	23.74	12.20	3.87	23.34	10.44
Black or African American	2,239	4.64	18.83	6.57	2.55	10.69	3.63
White	12,863	1.26	14.39	7.67	1.41	13.63	6.49
Other	1,085	5.81	23.95	12.28	8.11	39.57	15.62
White & Black or African American	157	2.55	22.56	5.78	0.64	0.66	0.22
White & Hispanic or Latino	625	3.04	20.33	11.91	1.60	21.21	10.75
White & Other	478	4.60	31.18	13.74	4.39	21.89	8.02
Black or African American & Hispanic or Latino	68	11.76	72.54	36.17	17.65	62.88	19.68
Black or African American & Other	60	0.00	0.00	0.00	0.00	0.00	0.00
Hispanic or Latino & Other	102	4.90	4.95	1.12	0.98	2.56	0.65
Pair Gender							
Male, Male	4,346	2.81	13.50	4.98	4.00	19.25	5.88
Female, Female	4,618	2.23	9.76	1.68	1.47	7.69	1.44
Male, Female	11,451	2.20	21.09	11.91	2.05	18.75	9.58
Household Size							
Two	5,555	0.40	1.30	0.27	0.58	1.97	0.47
Three	5,509	2.40	24.25	12.45	2.52	21.92	9.70
Four or More	9,351	3.45	22.00	10.95	3.27	22.27	10.10

Table J.2 2005 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	n	Before res.pr.nr[1] (SDUWT*PR05WT10*PR05WT11) % Unweighted	% Weighted[2]	% Outwinsor[3]	After res.pr.nr[1] (SDUWT*PR05WT10*...*PR05WT12) % Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region							
Northeast	4,094	2.42	21.03	10.83	2.39	22.47	9.96
South	6,074	1.94	12.80	5.04	1.88	12.54	4.80
Midwest	5,812	2.62	16.38	7.15	2.48	14.47	5.74
West	4,435	2.44	22.84	13.75	2.73	20.79	10.83
Quarter							
Quarter 1	4,956	2.14	21.42	12.58	1.90	20.06	11.35
Quarter 2	5,369	2.44	13.81	5.06	2.96	15.11	4.92
Quarter 3	5,063	2.11	19.35	10.75	1.88	15.69	7.79
Quarter 4	5,027	2.65	15.89	6.62	2.57	16.91	6.06
% Hispanic or Latino in Segment							
50-100%	1,641	3.47	34.74	21.64	3.72	33.81	17.94
10-50%	3,633	2.45	13.82	5.19	3.63	14.97	5.25
<10%	15,141	2.19	15.67	7.51	1.88	14.66	6.43
% Black or African American in Segment							
50-100%	1,557	5.52	18.38	4.03	4.05	13.14	2.47
10-50%	3,156	2.66	20.24	9.08	2.66	18.93	8.15
<10%	15,702	1.96	16.99	9.25	2.10	16.86	7.87
% Owner-Occupied DUs[1] in Segment							
50-100%	15,570	2.17	16.86	8.47	2.20	16.62	7.48
10-50%	3,780	3.44	22.27	10.97	3.23	19.27	8.17
<10%	1,065	0.85	8.26	2.28	1.22	7.45	1.91
Combined Median Rent/Housing Value							
1st Quintile	3,456	2.60	17.14	6.53	1.82	13.43	4.42
2nd Quintile	4,646	2.54	17.28	7.54	1.72	15.02	6.16
3rd Quintile	4,480	1.88	18.42	10.38	2.21	16.00	8.02
4th Quintile	4,283	2.47	19.67	11.25	3.43	19.68	9.75
5th Quintile	3,550	2.23	15.48	7.54	2.48	19.02	7.95
Population Density							
Large MSA[1]	8,507	3.00	21.78	12.20	3.24	22.41	10.61
Medium to Small MSA[1]	10,076	1.77	13.04	5.30	1.84	10.84	4.12
Non-MSA,[1] Urban	536	2.43	9.63	1.46	0.93	6.06	1.40
Non-MSA,[1] Rural	1,296	2.39	13.59	3.12	0.85	6.44	1.03
Group Quarters							
Group	429	3.73	13.59	1.90	1.17	4.60	0.79
Nongroup	19,986	2.31	17.69	8.85	2.36	16.98	7.53

[1] This step used demographic variables from screener data for all responding person pairs; DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PR = pair, Res = respondent, SDU = screener dwelling unit.
[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 nonextreme 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 2005 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12)			After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	20,415	2.26	12.19	4.00	1.93	9.47	1.91	1.00	6.42	0.89
Pair Age Group										
12-17, 12-17	3,913	0.54	2.82	0.55	0.72	2.83	0.55	0.20	1.03	0.27
12-17, 18-25	2,643	1.59	6.14	1.39	1.40	6.23	1.40	0.72	3.38	0.87
12-17, 26-34	724	1.38	4.63	1.24	1.66	5.86	1.23	0.69	2.46	0.78
12-17, 35-49	3,288	0.91	2.83	0.76	0.76	2.63	0.52	0.40	1.35	0.27
12-17, 50+	453	0.44	1.31	0.16	0.00	0.00	0.00	0.00	0.00	0.00
18-25, 18-25	4,255	3.48	15.68	4.53	2.89	11.87	2.10	1.34	5.56	1.07
18-25, 26-34	900	3.67	13.68	5.30	3.78	10.90	1.54	1.89	4.38	0.50
18-25, 35-49	1,119	4.92	12.83	2.51	3.13	7.36	0.86	1.07	2.33	0.10
18-25, 50+	509	3.54	11.37	2.53	1.57	4.35	1.01	1.18	3.86	0.62
26-34, 26-34	688	3.49	14.64	4.28	1.31	5.63	0.98	0.58	0.69	0.05
26-34, 35-49	382	6.02	27.48	11.28	2.88	8.64	1.63	1.83	4.82	0.40
26-34, 50+	138	2.90	9.47	1.18	0.72	3.76	0.36	0.00	0.00	0.00
35-49, 35-49	565	4.07	22.82	13.38	5.66	18.71	4.31	3.89	15.29	2.03
35-49, 50+	254	3.15	26.48	5.47	4.33	23.10	5.81	3.15	17.21	2.75
50+, 50+	584	3.60	10.44	3.29	4.97	15.06	2.82	4.62	13.58	1.56
Pair Race/Ethnicity										
Hispanic or Latino	2,785	3.63	17.53	7.35	2.30	11.44	2.13	1.22	6.22	1.42
Black or African American	2,195	2.60	8.41	1.34	2.32	8.75	1.31	1.23	4.91	0.47
White	12,582	1.34	8.15	2.75	1.34	7.83	1.84	0.61	5.66	0.71
Other	1,054	8.06	36.27	11.92	7.21	25.25	4.15	4.93	20.56	2.53
White & Black or African American	171	1.75	5.90	1.95	0.58	0.62	0.20	0.58	3.78	0.07
White & Hispanic or Latino	644	1.24	4.84	0.81	2.64	5.75	0.61	0.31	0.64	0.20
White & Other	645	3.41	10.13	1.59	1.86	6.08	0.79	0.93	2.78	0.19
Black or African American & Hispanic or Latino	82	14.63	42.24	15.26	7.32	20.39	5.80	6.10	18.12	5.01
Black or African American & Other	128	1.56	36.55	8.24	0.00	0.00	0.00	0.00	0.00	0.00
Hispanic or Latino & Other	129	3.10	40.00	7.79	0.00	0.00	0.00	0.78	1.99	0.46
Pair Gender										
Male, Male	4,340	3.96	16.98	5.58	2.67	10.75	2.00	1.08	4.82	0.83
Female, Female	4,618	1.47	7.85	1.83	1.32	5.46	1.05	0.41	2.89	0.41
Male, Female	11,457	1.94	12.02	4.15	1.90	10.18	2.12	1.21	7.81	1.04
Household Size										
Two	5,555	0.58	1.95	0.47	0.56	2.33	0.38	0.29	1.17	0.19
Three	5,509	2.49	15.77	6.12	2.31	11.97	2.37	1.71	10.20	1.02
Four or More	9,351	3.13	15.76	4.78	2.53	11.96	2.49	1.02	7.22	1.20

Table J.3 2005 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	n	Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12)			After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	4,094	2.27	14.04	3.49	2.17	13.87	4.35	1.29	9.56	1.90
South	6,074	1.78	7.07	2.15	1.55	5.31	0.97	0.79	2.80	0.43
Midwest	5,812	2.41	10.90	3.63	1.98	10.14	1.38	1.22	7.71	0.63
West	4,435	2.73	18.86	7.25	2.19	11.29	1.82	0.74	7.91	0.99
Quarter										
Quarter 1	4,956	1.82	15.02	6.49	2.00	13.71	3.66	1.29	9.77	1.62
Quarter 2	5,369	2.94	11.92	2.54	1.97	5.95	0.95	0.73	3.34	0.36
Quarter 3	5,063	1.80	12.66	5.26	2.21	9.64	1.91	1.19	6.76	1.01
Quarter 4	5,027	2.45	9.21	1.81	1.55	8.70	1.17	0.84	5.90	0.60
% Hispanic or Latino in Segment										
50-100%	1,641	3.29	21.78	9.26	2.80	11.83	2.00	1.04	5.91	1.14
10-50%	3,633	3.72	12.21	2.40	2.37	10.70	2.42	1.21	6.63	1.00
<10%	15,141	1.80	10.55	3.61	1.74	8.68	1.74	0.95	6.44	0.82
% Black or African American in Segment										
50-100%	1,557	3.92	13.19	2.55	3.47	11.70	2.11	1.67	6.80	1.12
10-50%	3,156	2.53	11.36	2.47	2.00	10.35	2.45	1.20	7.74	1.20
<10%	15,702	2.04	12.27	4.50	1.77	9.04	1.77	0.90	6.08	0.80
% Owner-Occupied DUs[1] in Segment										
50-100%	15,570	2.11	11.49	4.00	1.73	8.94	1.87	0.90	6.45	0.87
10-50%	3,780	3.15	15.75	4.18	2.86	12.25	2.16	1.48	6.57	1.00
<10%	1,065	1.41	9.07	2.74	1.60	6.10	1.35	0.85	3.37	0.84
Combined Median Rent/Housing Value										
1st Quintile	3,456	1.59	5.52	1.61	1.65	5.86	1.10	0.84	3.33	0.49
2nd Quintile	4,646	1.68	8.54	2.05	1.76	6.14	1.01	0.86	3.24	0.32
3rd Quintile	4,480	2.14	12.49	4.75	1.50	12.26	2.31	0.96	9.31	1.32
4th Quintile	4,283	3.34	15.91	4.87	2.73	10.79	2.96	1.21	6.90	1.25
5th Quintile	3,550	2.54	15.88	5.82	2.03	10.91	1.80	1.15	8.17	0.90
Population Density										
Large MSA[1]	8,507	3.17	17.92	6.34	2.59	13.38	2.87	1.48	9.58	1.42
Medium to Small MSA[1]	10,076	1.75	5.28	1.23	1.54	4.27	0.62	0.60	1.90	0.15
Non-MSA,[1] Urban	536	0.93	6.06	1.49	1.49	3.85	1.17	0.75	3.42	0.97
Non-MSA,[1] Rural	1,296	0.85	3.82	0.28	0.93	8.26	1.53	1.16	7.49	0.80
Group Quarters										
Group	429	0.93	4.68	0.78	1.40	6.46	1.00	0.47	3.21	1.22
Nongroup	19,986	2.29	12.22	4.02	1.95	9.48	1.91	1.02	6.43	0.89

Table J.3 2005 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12)			After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13)			Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
***Pair Relationship Domain*[4]**										
Parent-Child (12-14)	2,237	1.12	2.86	0.74	1.25	3.80	0.57	0.54	1.35	0.29
Parent-Child (12-17)	4,111	0.92	2.54	0.72	0.97	2.78	0.57	0.49	1.40	0.31
Parent-Child (12-20)	4,883	1.60	5.74	1.33	1.33	3.75	0.66	0.61	1.67	0.29
Sibling (12-14)-Sibling (15-17)	2,279	0.53	2.39	0.51	0.66	2.73	0.58	0.18	1.01	0.31
Sibling (12-17)-Sibling (18-25)	2,340	1.50	6.17	1.49	1.15	5.82	1.43	0.73	3.33	0.90
Spouse-Spouse/Partner-Partner	4,263	1.88	14.59	6.08	1.85	13.24	3.09	1.60	11.65	1.64
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	2,038	1.91	23.40	9.12	2.80	20.51	5.33	2.11	16.71	2.59

[1] This step used demographic variables from questionnaire data for all responding person pairs; DU = dwelling unit, EV = extreme value adjustment, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, Res = respondent, SDU = screener dwelling unit.

[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 nonextreme 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.

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

Table K.1 2005 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	20,415	203,624,333	203,624,333	203,624,333	0.00	-0.00
Pair Age Group						
12-17, 12-17	3,913	7,778,367	7,750,911	7,750,911	0.35	-0.00
12-17, 18-25	2,643	7,753,424	7,793,483	7,793,483	-0.51	-0.00
12-17, 26-34	724	5,232,222	5,228,354	5,228,354	0.07	-0.00
12-17, 35-49	3,288	32,140,587	32,176,810	32,176,810	-0.11	-0.00
12-17, 50+	453	9,721,960	9,716,846	9,716,846	0.05	0.00
18-25, 18-25	4,255	11,630,707	11,555,084	11,555,084	0.65	-0.00
18-25, 26-34	900	7,142,436	7,038,115	7,038,115	1.48	0.00
18-25, 35-49	1,119	17,066,849	16,965,106	16,965,106	0.60	0.00
18-25, 50+	509	13,038,038	13,309,188	13,309,188	-2.04	-0.00
26-34, 26-34	688	9,476,455	9,786,266	9,786,266	-3.17	0.00
26-34, 35-49	382	8,731,040	8,459,157	8,459,157	3.21	0.00
26-34, 50+	138	8,272,601	8,106,363	8,106,363	2.05	0.00
35-49, 35-49	565	20,153,241	20,366,951	20,366,951	-1.05	0.00
35-49, 50+	254	15,251,830	15,229,359	15,229,359	0.15	0.00
50+, 50+	584	30,234,576	30,142,341	30,142,341	0.31	0.00
Pair Race/Ethnicity						
Hispanic or Latino	2,785	32,375,082	31,295,413	31,295,413	3.45	-0.00
Black or African American	2,195	21,426,825	22,984,781	22,984,781	-6.78	-0.00
White	12,582	120,278,830	122,349,486	122,349,486	-1.69	0.00
Other	1,054	13,047,727	13,136,877	13,136,877	-0.68	0.00
White & Black or African American	171	1,691,891	1,472,612	1,472,612	14.89	0.00
White & Hispanic or Latino	644	5,098,345	6,157,064	6,157,064	-17.20	-0.00
White & Other	645	4,623,148	4,125,317	4,125,317	12.07	0.00
Black or African American & Hispanic or Latino	82	949,975	818,625	818,625	16.05	-0.00
Black or African American & Other	128	2,135,934	557,545	557,545	283.10	0.00
Hispanic or Latino & Other	129	1,996,575	726,612	726,612	174.78	-0.00
Pair Gender						
Male, Male	4,340	36,588,422	36,777,458	36,777,458	-0.51	-0.00
Female, Female	4,618	35,424,584	35,362,304	35,362,304	0.18	-0.00
Male, Female	11,457	131,611,327	131,484,571	131,484,571	0.10	-0.00
Pair Relationship Domain[3,4,5]						
Parent-Child (12-14)*	2,237	11,876,915	12,605,305	12,605,305	-5.78	-0.00
Parent-Child (12-17)*	4,111	23,837,086	24,989,554	24,989,554	-4.61	-0.00
Parent-Child (15-17)*	1,874	11,960,171	12,384,249	12,384,249	-3.42	-0.00
Parent-Child (12-20)*	4,883	32,099,074	33,243,054	33,243,054	-3.44	-0.00
Parent*-Child (12-14)	2,237	18,475,975	19,341,546	19,341,546	-4.48	-0.00
Parent*-Child (12-17)	4,111	30,827,786	31,794,138	31,794,138	-3.04	-0.00
Parent*-Child (15-17)	1,874	18,981,841	19,266,277	18,891,276	0.48	1.99
Parent*-Child (12-20)	4,883	38,280,978	38,785,417	38,785,417	-1.30	-0.00
Sibling (12-14)-Sibling (15-17)*	2,279	4,119,480	4,105,965	4,105,965	0.33	-0.00
Sibling (12-17)-Sibling (18-25)*	2,340	5,597,794	5,470,081	5,470,081	2.33	-0.00
Spouse-Spouse/Partner-Partner	4,263	68,935,916	67,618,828	67,618,828	1.95	-0.00
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	2,038	25,276,542	29,816,760	29,816,760	-15.23	-0.00

Table K.1 2005 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	5,555	52,777,497	52,777,497	52,777,497	-0.00	-0.00
Three	5,509	52,170,053	52,170,053	52,170,053	0.00	-0.00
Four or More	9,351	98,676,783	98,676,783	98,676,783	0.00	-0.00
Census Region						
Northeast	4,094	38,550,531	38,550,531	38,550,531	-0.00	-0.00
South	6,074	70,350,650	70,350,650	70,350,650	-0.00	-0.00
Midwest	5,812	43,250,743	43,250,743	43,250,743	0.00	0.00
West	4,435	51,472,408	51,472,408	51,472,408	0.00	0.00
Quarter						
Quarter 1	4,956	49,824,197	49,824,197	49,824,197	-0.00	-0.00
Quarter 2	5,369	51,544,056	51,544,056	51,544,056	-0.00	-0.00
Quarter 3	5,063	51,209,673	51,209,673	51,209,673	0.00	-0.00
Quarter 4	5,027	51,046,407	51,046,407	51,046,407	0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	1,641	23,323,051	23,323,051	23,323,051	-0.00	-0.00
10-50%	3,633	42,728,034	42,728,034	42,728,034	0.00	-0.00
<10%	15,141	137,573,248	137,573,248	137,573,248	0.00	-0.00
% Black or African American in Segment						
50-100%	1,557	15,847,421	15,847,421	15,847,421	-0.00	-0.00
10-50%	3,156	34,318,207	34,318,207	34,318,207	-0.00	-0.00
<10%	15,702	153,458,705	153,458,705	153,458,705	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	15,570	164,748,981	164,748,981	164,748,981	0.00	-0.00
10-50%	3,780	35,351,772	35,351,772	35,351,772	-0.00	-0.00
<10%	1,065	3,523,579	3,523,579	3,523,579	-0.00	-0.00
Combined Median						
Rent/Housing Value						
1st Quintile	3,456	29,411,320	29,411,320	29,411,320	-0.00	-0.00
2nd Quintile	4,646	41,442,823	41,442,823	41,442,823	0.00	-0.00
3rd Quintile	4,480	42,664,619	42,664,619	42,664,619	0.00	-0.00
4th Quintile	4,283	46,471,411	46,471,411	46,471,411	-0.00	-0.00
5th Quintile	3,550	43,634,160	43,634,159	43,634,159	0.00	-0.00
Population Density						
Large MSA	8,507	112,132,120	112,132,120	112,132,120	0.00	-0.00
Medium to Small MSA	10,076	78,242,518	78,242,518	78,242,518	0.00	-0.00
Non-MSA, Urban	536	3,641,652	3,641,652	3,641,652	0.00	-0.00
Non-MSA, Rural	1,296	9,608,043	9,608,043	9,608,043	0.00	0.00
Group Quarters						
Group	429	986,965	986,965	986,965	0.00	-0.00
Nongroup	19,986	202,637,368	202,637,368	202,637,368	0.00	-0.00

DU = dwelling unit, MSA = metropolitan statistical area, SDU = screener dwelling unit.

[1] YR05WT1*...*YR05WT9*PR05WT10*...*PR05WT12 (before person pair poststratification).

[2] YR05WT1*...*YR05WT9*PR05WT10*...*PR05WT13 (after person pair poststratification).

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

[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

Table L.1 2005 NSDUH Selected Pair-Level Weight Summary Statistics

		SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)						Before sel.pr.ps[1] (SDUWT*PR05WT10)						After sel.pr.ps[1] (SDUWT*PR05WT10*PR05WT11)					
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]**
Total	26,562	17	488	740	1,192	9,209	1.47	33	1,090	2,816	7,576	2,655,517	22.15	11	1,044	2,757	7,395	1,195,500	9.87
Pair Age Group																			
12-17, 12-17	4,476	21	428	676	1,082	6,971	1.53	33	664	1,107	2,101	67,246	2.75	11	550	1,092	2,106	19,315	2.36
12-17, 18-25	3,283	37	523	778	1,211	5,298	1.42	54	827	1,402	2,684	90,491	3.28	15	880	1,579	3,033	19,041	2.00
12-17, 26-34	879	25	459	707	1,100	9,209	1.58	101	2,230	3,728	6,659	46,899	2.14	77	2,008	3,695	7,056	63,874	2.45
12-17, 35-49	4,187	22	453	719	1,157	6,709	1.46	109	3,078	5,472	9,351	122,773	2.26	74	2,416	4,851	9,594	101,837	2.36
12-17, 50+	623	45	462	717	1,188	2,938	1.44	578	6,180	10,180	17,466	180,407	2.34	234	5,635	10,229	17,710	124,069	2.22
18-25, 18-25	5,508	22	511	758	1,216	6,540	1.46	41	787	1,240	2,329	99,606	3.35	15	612	1,201	2,562	24,051	2.50
18-25, 26-34	1,206	17	509	757	1,229	4,838	1.45	95	2,483	3,931	6,567	93,052	2.90	43	2,039	3,449	6,628	65,356	2.75
18-25, 35-49	1,609	26	520	776	1,290	4,614	1.44	229	3,521	6,477	11,758	127,316	2.39	92	2,858	6,055	13,112	96,087	2.42
18-25, 50+	763	57	587	824	1,318	5,120	1.38	684	7,390	11,549	18,886	269,694	2.55	250	6,513	11,502	21,271	153,062	2.19
26-34, 26-34	957	18	544	756	1,307	3,579	1.45	133	4,593	7,199	12,486	111,292	2.19	241	3,714	6,496	10,842	216,449	3.46
26-34, 35-49	563	53	451	701	1,241	3,866	1.52	624	4,736	8,078	14,030	2,655,517	44.55	419	4,029	7,956	14,900	514,944	5.64
26-34, 50+	220	55	589	810	1,369	4,390	1.39	1,104	11,717	19,528	32,179	1,131,711	8.04	426	10,692	24,477	46,704	451,062	2.49
35-49, 35-49	867	49	502	750	1,199	4,243	1.45	621	5,681	9,285	15,081	2,554,945	25.65	228	5,201	9,968	17,437	1,195,500	11.61
35-49, 50+	447	46	525	774	1,245	4,518	1.45	893	9,517	16,278	28,566	915,012	6.50	676	8,155	15,417	32,820	799,764	5.54
50+, 50+	974	67	512	740	1,208	2,883	1.40	1,939	13,507	20,319	32,748	2,023,043	8.27	534	13,246	23,294	37,614	1,178,714	3.20
Pair Race/Ethnicity																			
Hispanic or Latino	3,587	18	505	851	1,378	6,971	1.45	45	1,239	3,146	8,064	2,023,043	23.61	11	1,111	3,117	8,051	1,178,714	11.43
Black or African American	2,706	42	608	893	1,282	5,678	1.41	80	1,252	3,225	8,031	915,012	9.27	23	1,256	3,310	8,016	747,920	7.03
White	16,779	28	502	714	1,111	5,298	1.41	45	1,066	2,741	7,402	2,554,945	19.29	20	1,041	2,678	7,159	1,195,500	10.03
Other	1,527	23	210	602	1,270	9,209	1.85	33	760	2,198	7,311	436,093	8.18	13	845	2,549	7,726	615,025	9.19
White & Black or African American	195	25	522	877	1,315	4,614	1.50	47	1,302	3,793	9,439	622,357	19.38	54	1,194	2,920	7,317	136,497	5.04
White & Hispanic or Latino	835	17	464	796	1,301	5,120	1.49	41	1,258	3,180	8,194	696,578	11.57	38	1,001	2,666	7,664	515,824	9.96
White & Other	640	23	209	566	1,026	6,709	1.91	70	835	2,326	7,470	579,993	14.11	26	787	2,175	6,706	415,856	9.73
Black or African American & Hispanic or Latino	82	27	486	820	1,213	3,579	1.60	65	827	1,783	6,690	2,655,517	45.77	15	517	1,763	5,237	265,448	10.55
Black or African American & Other	81	58	451	747	1,118	2,981	1.50	176	1,379	3,604	8,307	98,733	3.82	102	655	1,584	4,597	70,916	5.78
Hispanic or Latino & Other	130	37	184	453	1,028	4,518	2.13	47	672	1,786	5,233	317,289	16.12	16	727	1,826	5,365	83,082	4.91

Table L.1 2005 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

		SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9)						Before sel.pr.ps[1] (SDUWT*PR05WT10)						After sel.pr.ps[1] (SDUWT*PR05WT10*PR05WT11)					
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 Gender																			
Male, Male	5,769	21	490	763	1,235	6,540	1.48	33	1,065	2,522	6,458	1,131,711	14.03	21	1,002	2,481	6,210	514,944	6.40
Female, Female	5,629	25	484	723	1,173	9,209	1.48	45	1,024	2,582	6,493	167,463	4.02	20	958	2,435	6,196	168,897	4.67
Male, Female	15,164	17	488	738	1,184	6,709	1.45	34	1,133	3,085	8,495	2,655,517	26.48	11	1,116	3,021	8,457	1,195,500	11.28
Household Size																			
Two	7,569	17	477	722	1,177	9,209	1.46	41	1,066	3,162	9,369	79,402	2.73	13	835	2,396	8,442	138,804	3.55
Three	7,194	22	492	737	1,163	4,674	1.43	33	1,197	3,201	6,659	2,655,517	52.45	11	1,175	3,162	6,652	1,195,500	13.27
Four or More	11,799	21	489	754	1,219	6,971	1.49	45	1,052	2,480	7,416	2,023,043	16.92	20	1,151	2,719	7,477	1,178,714	11.00
Census Region																			
Northeast	5,395	22	446	686	968	6,971	1.40	45	986	2,570	6,818	1,320,638	14.25	11	790	2,231	6,457	799,764	10.63
South	7,832	36	688	993	1,351	9,209	1.35	45	1,461	3,664	9,122	579,993	4.88	20	1,343	3,597	9,202	647,544	5.86
Midwest	7,546	23	514	639	811	5,540	1.32	33	922	2,492	6,302	2,554,945	34.18	15	1,028	2,409	5,949	987,576	10.82
West	5,789	17	290	722	1,677	6,540	1.64	39	1,017	2,789	8,428	2,655,517	34.74	38	1,022	2,907	8,407	1,195,500	13.31
Quarter																			
Quarter1	6,375	22	526	805	1,247	5,971	1.43	39	1,152	2,970	8,031	1,440,944	13.96	13	1,006	2,738	7,513	1,195,500	13.34
Quarter2	7,100	17	466	696	1,087	6,483	1.45	53	1,026	2,650	7,196	1,131,711	10.39	16	1,022	2,707	7,236	515,824	6.99
Quarter3	6,569	21	482	711	1,221	5,680	1.48	34	1,102	2,778	7,515	2,655,517	48.43	11	1,056	2,679	7,201	1,178,714	12.13
Quarter4	6,518	18	493	751	1,229	9,209	1.49	33	1,088	2,951	7,718	579,993	7.77	20	1,091	2,922	7,698	647,544	6.89
% Hispanic or Latino in Segment																			
50-100%	2,156	39	618	1,101	1,606	5,680	1.31	54	1,879	4,381	9,695	2,023,043	25.51	27	1,438	3,871	9,222	1,178,714	15.12
10-50%	4,813	17	569	934	1,452	6,971	1.39	49	1,409	3,331	9,074	696,578	6.80	11	1,198	3,338	9,092	549,314	6.11
<10%	19,593	18	435	694	1,060	9,209	1.48	33	999	2,580	7,026	2,655,517	25.81	13	984	2,559	6,880	1,195,500	9.60
% Black or African American in Segment																			
50-100%	1,955	36	583	829	1,148	6,971	1.46	45	1,192	3,154	7,894	215,515	3.93	20	1,104	3,099	7,825	204,624	4.61
10-50%	4,038	39	611	910	1,400	9,209	1.38	49	1,331	3,148	7,963	696,578	6.88	11	1,162	3,031	7,972	549,314	7.82
<10%	20,569	17	440	703	1,145	6,709	1.48	33	1,034	2,733	7,475	2,655,517	26.84	13	1,021	2,680	7,246	1,195,500	10.94
% Owner-Occupied DUs[1] in Segment																			
50-100%	20,312	18	477	723	1,152	6,971	1.46	33	1,114	2,973	7,817	2,655,517	23.10	16	1,153	3,019	7,920	1,195,500	9.60
10-50%	4,885	17	512	802	1,289	9,209	1.49	34	1,053	2,504	7,031	1,320,638	18.93	11	970	2,484	6,875	747,920	9.78
<10%	1,365	25	563	837	1,353	5,540	1.44	47	901	1,988	5,861	622,357	12.81	15	382	955	2,531	136,497	7.28
Combined Median Rent/Housing Value																			
1st Quintile	4,295	21	411	680	1,077	9,209	1.52	49	979	2,540	6,785	1,131,711	12.08	11	944	2,459	6,745	451,062	7.17
2nd Quintile	5,860	18	433	702	1,101	6,971	1.50	33	995	2,609	6,969	1,320,638	14.57	18	984	2,583	6,929	747,920	8.19
3rd Quintile	5,828	17	438	718	1,171	6,709	1.49	34	1,020	2,667	7,352	2,655,517	36.54	13	893	2,338	6,412	1,178,714	13.07
4th Quintile	5,742	23	503	758	1,240	6,540	1.44	51	1,178	2,956	7,951	2,554,945	27.09	15	1,126	3,053	7,960	987,576	10.53
5th Quintile	4,837	23	587	861	1,325	5,971	1.38	46	1,408	3,556	8,858	1,440,944	10.98	20	1,378	3,542	9,259	1,195,500	8.92

Table L.1 2005 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: YR05WT1*...*YR05WT9) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Before sel.pr.ps[1] (SDUWT*PR05WT10) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	After sel.pr.ps[1] (SDUWT*PR05WT10*PR05WT11) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Population Density																			
Large MSA[1]	11,496	36	654	925	1,416	6,971	1.33	45	1,528	3,908	9,516	2,554,945	20.17	11	1,416	3,697	9,503	1,195,500	10.41
Medium to Small MSA[1]	12,771	17	343	627	996	9,209	1.55	33	902	2,271	6,161	2,655,517	24.76	13	869	2,245	5,989	647,544	7.55
Non-MSA,[1] Urban	659	46	291	575	922	3,049	1.54	98	840	1,908	5,139	92,270	3.59	59	783	1,964	5,690	125,332	4.59
Non-MSA,[1] Rural	1,636	22	250	574	928	3,681	1.59	41	826	2,123	5,843	142,617	4.45	17	777	2,084	5,827	204,873	5.29
Group Quarters																			
Group	472	25	374	708	1,053	4,137	1.59	47	668	1,026	2,382	52,828	3.85	40	387	1,022	2,477	46,965	3.93
Nongroup	26,090	17	489	741	1,194	9,209	1.46	33	1,109	2,881	7,685	2,655,517	21.98	11	1,065	2,803	7,511	1,195,500	9.78

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

Table L.2 2005 NSDUH Respondent Pair-Level Weight Summary Statistics

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR05WT10*PR05WT11)						After res.pr.nr[1] (SDUWT*PR05WT10*...*PR05WT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,415	11	982	2,527	6,604	1,195,500	12.19	11	1,138	3,089	8,534	1,333,226	11.46
Pair Age Group													
12-17, 12-17	3,915	11	546	1,087	2,101	19,315	2.33	11	609	1,235	2,435	30,497	2.37
12-17, 18-25	2,658	15	876	1,589	3,075	19,041	2.01	15	1,006	1,933	3,740	30,475	2.10
12-17, 26-34	719	112	2,099	3,696	7,025	63,874	2.48	125	2,410	4,673	8,380	77,219	2.58
12-17, 35-49	3,293	74	2,424	4,848	9,588	101,837	2.36	76	2,893	5,801	11,920	114,644	2.44
12-17, 50+	450	234	5,295	10,061	16,837	124,069	2.24	259	7,298	13,427	25,227	171,585	2.26
18-25, 18-25	4,334	15	605	1,185	2,552	24,051	2.54	16	689	1,388	3,138	39,408	2.75
18-25, 26-34	847	43	2,055	3,431	6,440	65,356	2.74	44	2,427	4,640	9,317	151,393	2.99
18-25, 35-49	1,125	92	2,916	6,234	13,618	93,608	2.35	159	3,739	8,538	19,154	136,989	2.41
18-25, 50+	504	250	6,480	11,481	20,584	134,367	2.12	271	8,592	16,991	33,848	261,765	2.30
26-34, 26-34	667	241	3,638	6,421	10,638	216,449	3.63	244	4,503	7,935	14,788	324,290	4.21
26-34, 35-49	373	419	3,635	7,226	14,366	514,944	6.87	448	4,671	9,707	20,159	656,061	6.25
26-34, 50+	133	426	10,145	24,199	50,399	451,062	2.62	540	12,584	39,022	77,477	468,966	2.21
35-49, 35-49	565	228	5,210	9,866	17,437	1,195,500	12.54	238	6,771	13,344	24,500	1,229,676	10.74
35-49, 50+	253	676	8,300	15,938	32,348	799,764	6.49	1,115	11,245	24,925	53,730	1,333,226	5.46
50+, 50+	579	1,605	12,275	22,143	36,412	1,178,714	4.26	2,771	20,364	36,993	69,405	1,330,080	2.79
Pair Race/Ethnicity													
Hispanic or Latino	2,738	11	1,047	2,924	7,507	1,178,714	14.63	11	1,181	3,450	9,964	1,330,080	12.85
Black or African American	2,239	23	1,186	3,105	7,520	747,920	7.94	25	1,329	3,638	9,013	772,709	7.17
White	12,863	20	983	2,459	6,366	1,195,500	12.08	24	1,141	2,983	8,230	1,333,226	11.37
Other	1,085	13	691	2,048	6,015	615,025	14.37	13	840	2,787	9,313	1,102,727	14.79
White & Black or African American	157	54	1,311	3,145	7,317	136,497	5.38	54	1,416	3,471	8,653	161,153	5.76
White & Hispanic or Latino	625	38	924	2,701	7,218	515,824	12.23	38	1,027	3,043	8,521	705,131	13.15
White & Other	478	26	710	1,906	5,421	415,856	13.63	26	845	2,591	8,006	478,099	9.96
Black or African American & Hispanic or Latino	68	15	502	1,211	3,910	265,448	17.79	18	1,010	2,783	9,542	271,196	8.56
Black or African American & Other	60	102	358	1,657	4,483	70,916	6.77	103	893	2,698	6,576	121,000	6.20
Hispanic or Latino & Other	102	16	764	1,911	5,343	65,146	4.28	16	879	2,536	6,442	120,908	5.63
Pair Gender													
Male, Male	4,346	21	939	2,252	5,538	514,944	7.51	24	1,098	2,842	7,458	656,061	7.64
Female, Female	4,618	20	926	2,375	5,947	168,897	4.61	24	1,044	2,764	7,063	282,722	6.05
Male, Female	11,451	11	1,027	2,752	7,433	1,195,500	14.24	11	1,211	3,385	9,707	1,333,226	12.75
Household Size													
Two	5,555	13	781	2,101	7,013	104,467	3.72	13	893	2,458	9,186	233,856	4.89
Three	5,509	11	1,096	2,919	6,225	1,195,500	17.68	11	1,295	3,621	7,967	1,229,676	14.13
Four or More	9,351	20	1,102	2,542	6,777	1,178,714	12.68	24	1,278	3,107	8,690	1,333,226	13.28

Table L.2 2005 NSDUH Respondent Pair-Level Weight Summary Statistics (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PR05WT10*PR05WT11)						After res.pr.nr[1] (SDUWT*PR05WT10*...*PR05WT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region													
Northeast	4,094	11	678	1,977	5,544	799,764	13.68	11	775	2,364	7,368	1,333,226	14.46
South	6,074	20	1,275	3,367	8,329	647,544	6.38	24	1,466	4,009	10,282	822,683	7.31
Midwest	5,812	15	982	2,252	5,444	987,576	11.91	18	1,167	2,738	7,029	1,115,827	10.52
West	4,435	38	968	2,673	7,311	1,195,500	17.66	51	1,131	3,318	9,652	1,330,080	14.52
Quarter													
Quarter1	4,956	13	931	2,480	6,611	1,195,500	16.78	13	1,052	2,938	8,321	1,333,226	17.03
Quarter2	5,369	16	961	2,462	6,413	515,824	7.00	16	1,163	3,149	8,581	705,131	7.43
Quarter3	5,063	11	1,004	2,471	6,547	1,178,714	15.88	11	1,164	2,971	8,304	1,330,080	12.76
Quarter4	5,027	21	1,037	2,735	6,929	647,544	8.00	24	1,208	3,286	8,872	822,683	8.59
% Hispanic or Latino in Segment													
50-100%	1,641	27	1,325	3,522	8,026	1,178,714	18.85	27	1,485	4,164	10,313	1,330,080	16.55
10-50%	3,633	11	1,119	3,008	8,275	549,314	6.58	11	1,351	3,838	11,353	825,425	6.95
<10%	15,141	13	933	2,359	6,132	1,195,500	11.51	13	1,074	2,868	7,824	1,333,226	11.43
% Black or African American in Segment													
50-100%	1,557	23	1,046	2,885	7,025	175,759	4.37	25	1,279	3,608	9,051	263,005	4.99
10-50%	3,156	11	1,115	2,834	7,461	549,314	9.16	11	1,275	3,395	9,411	825,425	10.07
<10%	15,702	13	954	2,446	6,414	1,195,500	13.85	13	1,107	2,985	8,331	1,333,226	12.47
% Owner-Occupied DUs[1] in Segment													
50-100%	15,570	16	1,082	2,772	7,103	1,195,500	11.79	16	1,272	3,400	9,089	1,333,226	11.19
10-50%	3,780	11	923	2,297	6,211	747,920	12.51	11	1,056	2,770	7,804	772,709	11.06
<10%	1,065	15	379	883	2,387	136,497	8.45	16	436	1,100	2,917	161,153	8.24
Combined Median Rent/Housing Value													
1st Quintile	3,456	11	892	2,339	6,180	451,062	8.24	11	1,013	2,726	7,515	511,991	8.15
2nd Quintile	4,646	18	942	2,427	6,287	747,920	9.84	22	1,095	2,889	7,790	772,709	9.86
3rd Quintile	4,480	13	842	2,191	5,962	1,178,714	15.14	13	984	2,668	7,742	1,330,080	13.13
4th Quintile	4,283	15	1,056	2,729	6,868	987,576	14.27	15	1,234	3,434	9,228	1,333,226	13.18
5th Quintile	3,550	20	1,262	3,212	7,994	1,195,500	11.57	24	1,515	4,057	10,926	1,229,676	10.45
Population Density													
Large MSA[1]	8,507	11	1,318	3,383	8,490	1,195,500	13.53	11	1,573	4,274	11,397	1,333,226	11.72
Medium to Small MSA[1]	10,076	13	835	2,090	5,478	647,544	8.23	13	956	2,506	6,796	822,683	8.72
Non-MSA,[1] Urban	536	85	784	1,937	5,374	125,332	4.77	96	852	2,210	6,084	226,547	6.39
Non-MSA,[1] Rural	1,296	17	740	1,963	5,364	204,873	5.82	17	802	2,249	6,356	283,021	6.85
Group Quarters													
Group	429	40	397	993	2,448	46,965	4.10	40	414	1,184	2,649	60,854	4.59
Nongroup	19,986	11	1,002	2,584	6,731	1,195,500	12.07	11	1,168	3,173	8,709	1,333,226	11.32

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

Table L.3 2005 NSDUH Respondent Pair-Level Weight Summary Statistics

Domain	*n*	Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14) Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	20,415	11	1,138	3,089	8,534	1,333,226	11.46	3	1,038	3,030	8,504	1,420,149	9.82	3	1,022	3,016	8,485	1,088,808	9.26
Pair Age Group																			
12-17, 12-17	3,913	11	610	1,239	2,441	30,497	2.38	4	565	1,173	2,447	36,063	2.46	3	564	1,168	2,454	34,259	2.45
12-17, 18-25	2,643	15	1,006	1,929	3,740	30,475	2.12	3	905	1,892	3,819	35,834	2.22	3	898	1,888	3,855	33,150	2.18
12-17, 26-34	724	55	2,403	4,578	8,155	77,219	2.67	23	2,291	4,152	8,357	71,091	2.65	22	2,235	4,193	8,554	72,457	2.68
12-17, 35-49	3,288	76	2,894	5,835	11,932	114,644	2.42	18	2,812	5,796	11,952	133,678	2.50	17	2,775	5,700	11,972	131,065	2.50
12-17, 50+	453	259	7,159	13,156	24,932	171,585	2.28	181	6,456	12,828	25,986	173,061	2.36	163	6,439	12,721	26,255	171,383	2.38
18-25, 18-25	4,255	16	685	1,386	3,141	151,393	3.56	7	574	1,289	3,336	50,973	2.90	6	551	1,266	3,414	52,103	2.86
18-25, 26-34	900	44	2,205	4,297	8,826	227,571	3.64	28	1,692	3,956	8,636	116,569	3.39	24	1,635	3,855	8,655	115,789	3.40
18-25, 35-49	1,119	159	3,738	8,538	19,294	137,361	2.46	182	3,459	8,373	19,051	140,056	2.53	179	3,387	8,402	19,544	141,227	2.52
18-25, 50+	509	271	8,397	16,332	32,834	261,765	2.30	307	7,987	16,436	34,124	202,454	2.24	301	7,811	16,517	34,500	197,203	2.24
26-34, 26-34	688	244	4,361	7,721	14,470	324,290	4.12	134	3,789	7,465	15,009	347,809	4.32	111	3,696	7,260	14,940	353,290	4.28
26-34, 35-49	382	448	4,671	9,706	20,159	656,061	6.18	212	4,133	9,010	20,847	677,279	5.42	182	4,016	9,035	20,825	699,621	5.48
26-34, 50+	138	540	12,584	39,108	77,421	468,966	2.22	565	13,991	32,872	75,726	452,071	2.29	545	13,268	32,473	79,035	449,790	2.30
35-49, 35-49	565	238	6,706	13,184	24,079	1,229,676	10.94	96	6,791	14,746	30,903	1,142,697	7.70	77	6,486	14,699	31,354	962,248	7.20
35-49, 50+	254	1,115	10,503	24,772	55,349	1,333,226	5.47	995	8,900	21,708	59,057	1,420,149	5.28	919	8,788	21,414	62,842	1,088,808	4.51
50+, 50+	584	2,771	20,141	36,383	69,361	1,330,080	2.80	1,546	18,335	35,068	66,674	840,474	2.31	1,410	18,594	35,974	67,948	825,527	2.26
Pair Race/Ethnicity																			
Hispanic or Latino	2,785	11	1,211	3,583	10,025	1,330,080	12.68	8	1,106	3,423	9,651	840,474	9.11	8	1,090	3,431	9,755	825,527	9.11
Black or African American	2,195	25	1,328	3,592	8,681	263,005	5.08	16	1,233	3,528	9,175	259,131	5.52	15	1,205	3,521	9,244	263,301	5.58
White	12,582	24	1,141	2,983	8,246	1,333,226	11.46	12	1,086	3,013	8,334	1,420,149	10.63	11	1,072	3,020	8,353	1,088,808	9.67
Other	1,054	13	888	2,873	9,475	1,102,727	14.55	4	872	3,006	9,578	860,030	12.00	3	847	2,967	9,831	816,950	11.72
White & Black or African American	171	29	1,229	3,440	9,687	161,153	5.37	9	966	2,309	7,373	158,228	6.21	8	929	2,311	6,540	157,784	6.25
White & Hispanic or Latino	644	16	931	2,784	8,010	176,495	5.28	7	950	2,993	8,756	257,775	5.87	6	923	2,968	8,566	270,637	6.06
White & Other	645	54	864	2,480	6,747	171,045	4.53	23	573	1,892	5,857	243,828	6.47	22	564	1,866	5,762	260,286	6.86
Black or African American & Hispanic or Latino	82	18	1,043	2,692	10,612	90,879	4.50	8	591	1,984	6,499	118,927	5.63	7	572	2,158	6,300	122,756	5.70
Black or African American & Other	128	44	1,182	3,607	11,802	772,709	18.47	9	282	871	3,135	157,618	13.17	9	268	866	3,069	160,480	13.68
Hispanic or Latino & Other	129	15	804	2,499	6,667	705,131	24.15	3	254	742	2,136	329,284	33.01	3	247	745	1,943	329,957	33.18

Table L.3 2005 NSDUH Respondent Pair-Level Weight Summary Statistics (continued)

		Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12)						After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14)					
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 Gender																			
Male, Male	4,340	24	1,099	2,841	7,433	656,061	7.66	8	1,007	2,802	7,314	452,071	6.89	7	992	2,795	7,358	449,790	6.73
Female, Female	4,618	24	1,043	2,765	7,095	282,722	6.03	15	959	2,669	7,146	286,660	6.23	13	942	2,655	7,112	295,435	6.33
Male, Female	11,457	11	1,213	3,382	9,690	1,333,226	12.74	3	1,094	3,357	9,685	1,420,149	10.68	3	1,076	3,328	9,676	1,088,808	9.93
Household Size																			
Two	5,555	13	893	2,458	9,186	233,856	4.89	4	726	2,126	8,792	245,366	5.20	3	697	2,068	8,581	240,907	5.24
Three	5,509	11	1,295	3,621	7,967	1,229,676	14.13	3	1,258	3,640	8,125	681,352	8.80	3	1,238	3,620	8,108	650,172	8.71
Four or More	9,351	24	1,278	3,107	8,690	1,333,226	13.28	7	1,196	3,125	8,729	1,420,149	12.48	6	1,197	3,143	8,763	1,088,808	11.40
Census Region																			
Northeast	4,094	11	775	2,364	7,368	1,333,226	14.46	3	701	2,250	7,227	1,420,149	16.92	3	673	2,221	7,172	1,088,808	13.82
South	6,074	24	1,466	4,009	10,282	822,683	7.31	8	1,355	3,950	10,200	681,352	6.83	7	1,322	3,931	10,099	699,621	6.94
Midwest	5,812	18	1,167	2,738	7,029	1,115,827	10.52	19	1,056	2,680	6,921	781,631	9.44	19	1,052	2,694	6,869	749,182	9.38
West	4,435	51	1,131	3,318	9,652	1,330,080	14.52	36	1,034	3,241	9,795	860,030	8.72	34	1,019	3,220	9,940	825,527	8.56
Quarter																			
Quarter1	4,956	13	1,052	2,938	8,321	1,333,226	17.03	3	976	2,987	8,347	1,420,149	14.27	3	968	2,982	8,462	1,088,808	11.81
Quarter2	5,369	16	1,163	3,149	8,581	705,131	7.43	7	1,045	3,022	8,697	677,279	7.09	6	1,021	2,998	8,583	699,621	7.19
Quarter3	5,063	11	1,164	2,971	8,304	1,330,080	12.76	8	1,089	2,997	8,365	840,474	8.45	8	1,080	3,005	8,388	825,527	8.45
Quarter4	5,027	24	1,208	3,286	8,872	822,683	8.59	8	1,057	3,109	8,556	781,631	9.49	7	1,031	3,076	8,544	749,182	9.54
% Hispanic or Latino in Segment																			
50-100%	1,641	27	1,485	4,164	10,313	1,330,080	16.55	29	1,607	4,526	11,317	840,474	10.04	28	1,601	4,517	11,337	825,527	9.95
10-50%	3,633	11	1,351	3,838	11,353	825,425	6.95	8	1,155	3,584	10,469	1,142,697	7.96	8	1,140	3,552	10,483	962,248	7.30
<10%	15,141	13	1,074	2,868	7,824	1,333,226	11.43	3	977	2,789	7,731	1,420,149	10.10	3	963	2,766	7,728	1,088,808	9.47
% Black or African American in Segment																			
50-100%	1,557	25	1,279	3,608	9,051	263,005	4.99	9	1,057	3,316	8,861	259,131	5.52	9	1,045	3,303	8,880	257,397	5.56
10-50%	3,156	11	1,275	3,395	9,411	825,425	10.07	8	1,125	3,275	8,993	1,142,697	10.98	8	1,098	3,270	9,010	962,248	10.25
<10%	15,702	13	1,107	2,985	8,331	1,333,226	12.47	3	1,021	2,951	8,362	1,420,149	9.98	3	1,005	2,927	8,375	1,088,808	9.39
% Owner-Occupied DUs[1] in Segment																			
50-100%	15,570	16	1,272	3,400	9,089	1,333,226	11.19	5	1,180	3,371	9,092	1,420,149	9.83	4	1,164	3,336	9,113	1,088,808	9.17
10-50%	3,780	11	1,056	2,770	7,804	772,709	11.06	3	923	2,648	8,008	529,099	8.24	3	889	2,632	8,046	539,821	8.25
<10%	1,065	16	436	1,100	2,917	161,153	8.24	7	397	975	2,922	144,936	7.45	6	385	946	2,922	143,036	7.56
Combined Median Rent/Housing Value																			
1st Quintile	3,456	11	1,013	2,726	7,515	511,991	8.15	5	894	2,524	7,406	618,942	8.13	4	883	2,519	7,319	633,483	8.23
2nd Quintile	4,646	22	1,095	2,889	7,790	772,709	9.86	7	966	2,770	7,833	677,279	9.47	6	949	2,752	7,754	699,621	9.79
3rd Quintile	4,480	13	984	2,668	7,742	1,330,080	13.13	4	863	2,579	7,389	1,142,697	12.95	3	852	2,557	7,290	962,248	12.00
4th Quintile	4,283	15	1,234	3,434	9,228	1,333,226	13.18	3	1,171	3,547	9,556	1,420,149	10.37	3	1,149	3,530	9,645	1,088,808	8.77
5th Quintile	3,550	24	1,515	4,057	10,926	1,229,676	10.45	12	1,433	4,078	11,032	860,030	7.32	11	1,417	4,108	10,995	816,950	7.14

Table L.3 2005 NSDUH Respondent Pair-Level Weight Summary Statistics (continued)

		Before res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT12)						After res.pr.ps[1] (SDUWT*PR05WT10*...*PR05WT13)						Final Weight: After res.pr.ev[1] (SDUWT*PR05WT10*...*PR05WT14)					
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]**
Population Density																			
Large MSA[1]	8,507	11	1,573	4,274	11,397	1,333,226	11.72	8	1,515	4,247	11,615	1,420,149	9.23	8	1,501	4,288	11,645	1,088,808	8.36
Medium to Small MSA[1]	10,076	13	956	2,506	6,796	822,683	8.72	3	853	2,395	6,693	681,352	9.20	3	836	2,356	6,614	659,228	9.43
Non-MSA,[1] Urban	536	96	852	2,210	6,084	226,547	6.39	51	771	2,197	5,976	258,699	6.82	39	769	2,177	5,880	263,301	6.96
Non-MSA,[1] Rural	1,296	17	802	2,249	6,356	283,021	6.85	7	729	2,123	5,937	258,321	7.73	6	742	2,116	6,049	266,101	7.58
Group Quarters																			
Group	429	40	414	1,184	2,649	60,854	4.59	22	325	889	2,409	70,989	5.58	19	300	859	2,240	67,256	5.57
Nongroup	19,986	11	1,168	3,173	8,709	1,333,226	11.32	3	1,075	3,127	8,692	1,420,149	9.70	3	1,056	3,109	8,666	1,088,808	9.15
Pair Relationship Domain[4]																			
Parent-Child (12-14)	2,237	125	2,556	5,233	10,780	153,783	2.74	90	2,589	5,238	11,305	173,061	2.86	80	2,549	5,173	11,317	155,993	2.86
Parent-Child (12-17)	4,111	76	2,892	5,912	12,215	153,783	2.59	18	2,882	5,928	12,554	173,061	2.71	17	2,840	5,888	12,544	171,383	2.73
Parent-Child (12-20)	4,883	76	3,081	6,368	13,598	261,765	2.79	18	3,025	6,324	13,734	173,061	2.80	17	2,986	6,277	13,804	171,383	2.80
Sibling (12-14)-Sibling (15-17)	2,279	11	615	1,244	2,460	30,497	2.33	4	572	1,153	2,432	36,063	2.48	3	572	1,144	2,446	34,259	2.47
Sibling (12-17)-Sibling (18-25)	2,340	15	1,000	1,925	3,700	30,475	2.11	3	896	1,839	3,752	35,834	2.19	3	887	1,845	3,770	33,150	2.14
Spouse-Spouse/ Partner-Partner	4,263	16	1,087	3,726	13,342	1,333,226	14.11	7	897	3,479	13,668	1,420,149	11.91	6	872	3,465	13,859	1,088,808	10.84
Spouse-Spouse/ Partner-Partner with Children (Younger Than 18)	2,038	26	1,076	3,322	9,931	1,333,226	25.14	26	1,247	3,948	12,381	1,420,149	18.39	25	1,247	4,024	12,382	1,088,808	15.79

[1] This step used demographic variables from questionnaire data for all selected person pairs; DU = dwelling unit, EV = extreme value adjustment, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, Res = respondent, SDU = screener dwelling unit.

[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.

[3] Unequal weighting effect (UWE) is 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.

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."

For the 2005 National Survey on Drug Use and Health (NSDUH),[28] the imputation procedure used for most variables requiring imputation was the predictive mean neighborhood (PMN) method, which is a combination of predictive mean matching (Rubin, 1986) and unweighted random nearest neighbor hot deck (NNHD). No other type of hot-deck method was used to impute missing values in the 2005 survey. Although only one hot-deck imputation method was used in the 2005 survey, two other methods also have been used in past surveys. The three methods, which are each discussed in this report, are unweighted sequential hot deck, weighted sequential hot deck, and unweighted random NNHD. 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 was not used for other variables. In the 2000 survey, the unweighted sequential hot-deck method was used only for education and employment status and has not been used at all in the surveys since 2001. However, it remains in this appendix for historical purposes and for comparison with the other two methods. The third hot-deck method, weighted sequential hot deck, incorporated the sampling weights associated with each respondent. It was used in earlier surveys, but it was not used in the 2005 survey. More information on weighted sequential hot-deck imputation is available in Cox (1980, pp. 721-725) and Iannacchione (1982). The imputations of demographic and other variables unrelated to pair analyses are described in the 2005 NSDUH imputation report (Aldworth et al., 2007).

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 in Sections M.3 through 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

[28] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years 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 a resulting inconsistency with preexisting nonmissing values.

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 were 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 surveys to sort the files prior to imputation:

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

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

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

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

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

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

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. Any subsequent nonrespondent in the file received the stored donor response, which in turn resulted in a statistically imputed response. A starting value was needed if an item nonrespondent was the first record in a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value. Due to the fact that the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

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 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 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.4.1 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 surveys to sort the files prior to imputation: straight sort and serpentine sort. Both of these methods are described in detail in Section M.3.1.

M.4.2 Replacing Missing Values

The procedure used in the 2005 survey followed directly from Cox (1980). Specifically, once the imputation classes are formed, the data is divided into two datasets: one for respondents and one for nonrespondents. Scaled weights $v(j)$ are then derived for all nonrespondents using the following formula:

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

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

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

M.4.3 Benefits of Weighted Sequential Hot Deck

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there is the risk of several nonrespondents appearing next to one another in the sorted file. An imputed value could still be found for those cases, since the algorithm would select the previous respondent in the file. However, some modifications are required in the sorting procedure to prevent a single respondent from being the donor for several nonrespondents (see Section 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).

M.5 Unweighted Random Nearest Neighbor Hot Deck

As with the other methods, 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.5.1 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, and 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 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.

In application, if the predicted means were probabilities, the values of delta varied depending upon the value of the predicted mean. In this case, each delta was defined as 5 percent of the predicted probability if the probability was less than 0.5 and was defined as 5 percent of 1 minus the predicted probability if the probability was greater than 0.5. This allowed a looser delta for predicted probabilities close to 0.5 and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is shown in Table M.1.

Table M.1 Values of Delta for Various Predicted Probabilities

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

M.5.2 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[29] because weights were incorporated in

[29] In surveys prior to 2005, this was incorrectly calculated. Instead of each donor in the neighborhood (of size *n*) being assigned a probability of 1/*n* of being selected, the first and last donors in the neighborhood were assigned a probability of 1/(2(*n*-1)) of being selected, and the remaining donors were assigned a probability of 1/(*n*-1) of being selected. This was corrected in the 2005 survey.

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. If no donor pairs were available with predicted means within delta of the recipient donor's predicted mean(s), the neighborhood was abandoned and the donor with the closest predicted mean(s) was chosen.[30] This was done to reduce the potential for bias.

[30] In surveys prior to 2006, this sometimes was incorrectly applied. In some cases, the neighborhood was not sorted by the predicted means, and in other cases, a donor was randomly selected. There were also some situations where the donor with the closest predicted means was chosen with the delta constraint in place. These procedures will be corrected in the 2006 survey.

Appendix N: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

Appendix N: Univariate and Multivariate Predictive Mean Neighborhood Imputation Methods

N.1 Introduction

Since the introduction of the computer-assisted interviewing (CAI) method in 1999 for the National Survey on Drug Use and Health (NSDUH),[31] one imputation method has been used for most variables requiring imputation: predictive mean neighborhood (PMN). It was developed to cater to the specific needs of NSDUH. This approach has been used since the 1999 survey[32] and can be applied to one variable at a time or to several variables simultaneously. As described in this appendix, PMN incorporates predicted means from models and the assignment of imputed values using neighborhoods determined by those predicted means.

N.2 Overview

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

The PMN method is a combination of two commonly used imputation methods: a nonmodel-based hot deck (nearest neighbor) and a modification of the model-assisted predictive mean matching (PMM) method of Rubin (1986). The PMN method enhances the PMM method. Specifically, the PMN method can be applied to both discrete and continuous variables, either individually or jointly. The PMN method also enhances the nearest neighbor hot-deck (NNHD) method so 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, but 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 NSDUH, the set of covariates typically included demographic variables, as well as some other nonmissing pair-level variables. In the case of 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 variables 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.

---

[31] This report presents information from the 2005 National Survey on Drug Use and Health (NSDUH), an annual survey of the civilian, noninstitutionalized population of the United States aged 12 years or older. Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[32] After the 1999 survey, 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 predicted 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 predicted mean can be minimized by using suitable covariates.
- The PMM method is not a pure model-based method, because the predicted 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 predicted 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 the Predictive Mean Neighborhood Method

The PMN method is easily applicable to problems of both univariate and multivariate imputations. The need for univariate imputation arises when the value of a single 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 also can 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 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, because of the obvious problem of specifying the probability distribution underlying survey data, none of the existing solutions take the survey design into account. However, since the 1999 survey, in the application of the multivariate predictive mean neighborhood (MPMN) method to the imputation procedures, a multivariate model has been fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the hierarchy have a chance to be included in the covariate set for subsequent models in the hierarchy. In the multivariate modeling with MPMN, the innovative idea is to express the likelihood in the superpopulation model as a product of marginal and conditional likelihoods, which then allows for the use of univariate techniques for fitting multivariate (but conditional) predicted means.

In the application of person pair imputations, none of the variables imputed were part of a multivariate set, so 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 provided below is an abbreviated description of the method in the univariate case only. A description of the multivariate case is described in the 2005 NSDUH imputation report (Aldworth et al., 2007).

N.3 Outline and Description of Method

The procedure for implementing PMN in the 2005 survey, where imputed variables were not part of any multivariate set, entailed four 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. 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.3.3 Step 3: Computation of Predicted Means and Delta Neighborhoods

Once the model was fitted, the predicted means for item respondents and item nonrespondents were calculated using the model coefficients. This predicted 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 predicted mean of the item nonrespondent and the predicted means of every item respondent. Those item respondents whose predicted 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, *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 range were restricted to have had the same or similar marital status whenever possible. A small value of delta also could have been considered as a likeness constraint. Whenever insufficient donors were available to meet the likeness constraints, including the preset small value of delta, the constraints were loosened in priority order according to their perceived importance. As a last resort, if an insufficient number of donors was available to meet the logical constraints given the loosest set of likeness constraints allowable, a donor was found using a sequential hot deck, where matching was done on the predicted 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 3, 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.

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, making it easier to implement restrictions on the donor set.** Because the distance function is defined as a continuous function of the predicted 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 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 predicted means in the dataset via the Mahalanobis squared 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 predicted 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 make objective decisions 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 these rules, an "age match" occurs if the questionnaire age of one pair member matches a roster age in the other pair member's roster, and a "gender match" occurs if the questionnaire gender of one of the pair members matches a roster gender in the other pair member's roster. In the table below, 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 are 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 Table 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 that were not manifest in 2005 also are excluded from this table, provided the distribution of counts would have been unaffected by their exclusion from the code.

Table 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	17,103
2	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	2,154
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	154
4	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	242
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	33
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	8

Table 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	274
8	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	26
9	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	3
10	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	9
11	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	32
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 matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	1
14	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	32
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	4
16	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	1
17	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	170
18	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	26

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

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
19	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	7
20	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	4
21	Age matches exactly, gender off, 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	2
22	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	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
23	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	6
24	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	11
25	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	3
26	Age within two, gender matches exactly, exactly one MBRSEL in right place	Everything missing	1
27	Age and gender match exactly, MBRSEL missing for all roster members	Everything missing	1

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

Priority Condition	Rule		
	Pair Member A	**Pair Member B**	**Count**
28	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	28
29	Age within one, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	3
30	Age and gender match exactly, MBRSEL missing for all roster members	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	2
31	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	1
32	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	5
33	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender off, one MBRSEL, only two in household	1
34	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	10
35	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	2
36	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	33
37	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	7
38	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	3
39	Age and gender match exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	1
40	No match at all	No match at all	18

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 shown 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 provided in the columns "Constraint on A" and "Constraint on B." These constraints include age constraints, where a range of ages (e.g., 12 to 17) indicates that the value of the questionnaire edited age (AGE) is between the numbers shown. Also in this column, "child" and "children" are defined as (a) roster member(s) with nonmissing ages less than 18. The question FIPE3 asks is if the respondent is the parent of a selected 12- to 17-year-old. The responses provided 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.

Table 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					

Table 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						
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						

Table 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						
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				

Table 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 years older than B				11	2
	Missing	Grandparent						
	Grandparent	Missing	B at least 20 years 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
	Missing	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

Table 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/ other 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

Table 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 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 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 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 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

Table 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
52	Nonmissing, not 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 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 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 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 is 15-17, both sides have parents or spouses				5	3
	Sibling	Spouse						
55	Spouse	Sibling	One is 12-17, other is 18-25, both sides have parents or spouses				6	3
	Sibling	Spouse						

Table 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

Table 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						
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 older			13	3
	Unusual in-law code	Sibling	26 or older	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 older, 0 children, no bad data	15 or older, no spouse			23	4
	Not a child, parent, or sibling	Spouse/partner	15 or older, no spouse	15 or older, 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 Table 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 or older. 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, and thus no imputation was required. The first column in Table P.1 shows the hierarchical priority condition, with the frequency of occurrence for each priority condition in parentheses.

Table 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 (19,421)	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 than screener roster, equal to QD54A	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

Table 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 ≥ TGOODAGA		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	TGOODAGA ≤ TOTPEOPA (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	TGOODAGA ≤ SHHSIZE		TGOODAGA ≤ SHHSIZE	SHHSIZE
14 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1				TGOODAGA
15 (3)	Both missing or 1	Both missing or 1			SHHSIZE ≥ 2, SHHSIZE at least as large or larger than screener roster	SHHSIZE
16 (20)	Not equal, both > 1	TOTPEOP(B) = QD54 (B)	A: At least one age range variable less than min.[1] 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 (3)	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

Table 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 (820)	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 than 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 than screener roster, equal to QD54B	QD54B
19 (3)	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 than 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 than 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, TGOODAGA = QD54A	A: no B: no		QD54A
			TGOODAGA = TGOODAGB, 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

Table 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 (11)	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 (118)	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

Table 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] "Min." refers to the minimum possible within each age range based upon the ages of the two pair members.

P.2. Age variables

Table 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 or older 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 "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 or older. 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 used only if the age counts in each pair member's roster cannot conform to the minimum necessary or otherwise are not possible to incorporate. If after all edits the count for AGE1217 is missing, but the counts for other age groups are not missing, 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 that 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 who is identified as the other pair member with a perfect match on age and 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 counts for the pair members are unequal, then the count is set to missing. As with Table P.1, the first column in Table P.2 shows 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 provide details of the differences.

Table 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)[3]		AGE1217A < min. AGE1217B < min.		SHHSIZE = HHSIZE, SAGE1217 ≥ min.		SAGE1217
4 (7)[4]		AGE1217A = AGE1217B, both ≥ min.	Another count except 12-17 < min.			AGE1217A
5 (0)[5]		AGE1217A not equal to AGE1217B, both ≥ min.	AGE1825A < min., AGE1825B ≥ min.			AGE1217B
6 (0)[6]			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	AGE1217B
9 (1)[7]					A & B: none missing A has better match measure than B	AGE1217A

Table 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 (1)[8]	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 (1)[9]					A & B: none missing Age (A) ≥ Age (B)	AGE1217A
					A & B: none missing Age (B) > Age (A)	AGE1217B
12 (0)[10]						missing
13 (18,624)		AGE1217A = AGE1217B	All other counts equal across pair members			AGE1217A
14 (471)		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 (54)					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 (81)					A & B: none missing Age (A) ≥ Age (B)	AGE1217A
					A & B: none missing Age (B) > Age (A)	AGE1217B

Table 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

Table 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

Table 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 (529)		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 $\geq$ min.		SAGE-1217
41 (0)		AGE1217A = AGE1217B, both $\geq$ min.				AGE1217A
42 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B $\geq$ min.			AGE1217B
			AGE1825B < min. AGE1825A $\geq$ min.			AGE1217A
43 (0)					Fewer roster entries missing in B than A	AGE1217B
					Fewer roster entries missing in A than B	AGE1217A

Table 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 (399)		Priority conditions 39-47 not met				AGE1217B

Table 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 (110)	TGOODAGA = HHSIZE					TAG1217A
	TGOODAGB = HHSIZE					TAG1217B
50 (102)[11]	SHHSIZE = HHSIZE	AGE1217A, AGE1217B ≤ SAGE1217		AGE1217A & B ≤ SAGE1217		SAGE1217
51 (0)[12]	SHHSIZE = HHSIZE, HHSIZE = EXC1217	AGE1217 missing	Other counts not missing, AGE011A equals AGE011B			MIN1217
52 (18)[13]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, either the two TOTPEOP's > 0, or SHHSIZE = HHSIZE			AGE1217A equals SAGE1217		AGE1217A
				AGE1217B equals SAGE1217		AGE1217B

Table 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:
53 (4)	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, SHHSIZE = HHSIZE	AGE1217 missing	At least 3 of the other counts missing			SAGE1217

[1] The frequency of priority condition #1 for AGE3549 and AGE50p was 0.
[2] The frequency of priority condition #2 for AGE1825 and AGE2634 was 1. The frequency for AGE3549 was 2.
[3] The frequency of priority condition #3 for AGE1825 was 1.
[4] The following frequencies were observed for priority condition #4: AGE1220 and AGE1825, 4; AGE1214, AGE2634, and AGE3549, 8; and AGE50p, 10.
[5] The frequency of priority condition #5 for AGE1825 and AGE2634 was 2.
[6] The frequency of priority condition #6 for AGE3549 and AGE50p was 1.
[7] The frequency of priority condition #9 for AGE1220 and AGE1825 was 2. The frequency for AGE1214, AGE2634, AGE3549, and AGE50p was 0.
[8] The frequency of priority condition #10 for AGE1214, AGE2634, AGE3549, and AGE50p was 0.
[9] The frequency of priority condition #11 for AGE1825, AGE2634, AGE3549, and AGE50p was 0. The frequency for AGE1220 was 2.
[10] The frequency of priority condition #12 for AGE1220 was 1.
[11] The following frequencies were observed for priority condition #50: AGE1220, 99; AGE1825, 96; AGE2634 and AGE3549, 101; and AGE50p, 103.
[12] The frequency of priority condition #51 for the age range counts other than AGE1825 was 4. The frequency for AGE2634 and AGE3549 was 1.
[13] The following frequencies were observed for priority condition #52: AGE1214 and AGE1825, 17; AGE1220, 15; AGE3549, 16; and AGE50p, 21.

Appendix Q: Pair Relationship, Multiplicity, and Household Count Model Summaries

Appendix Q: Pair Relationship, Multiplicity, and Household Count Model Summaries

Q.1 Introduction

The tables 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 predicted 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. Thus, separate tables 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. These models were applied at a pair level, whereas some of the variables in the models were applied 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 predicted means in the imputation models. Descriptions of questionnaire-derived variables are described in detail in Chapter 6 of 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 used often as covariates in both types of models for the PMN procedures.

Census Region

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

MSA

The metropolitan statistical area (MSA) 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 that the respondents lived in according to modified 2000 census data, which was adjusted to more recent data from Claritas Inc.[33] This variable had three levels:

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

segment in MSA with 1 million or more persons; segment in MSA with fewer than 250,000 persons or 250,000 to 999,999 persons; and segment not in MSA.

Categorical Percent Hispanic or Latino in Segment

The categorical percent Hispanic or Latino in segment variable was used to categorize segments according to the concentration of Hispanics or Latinos in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 20 percent, 20 to 70 percent, and more than 70 percent.

Categorical Percent Owner-Occupied Households in Segment

The categorical percent owner-occupied households in segment 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 2000 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.

Categorical Percent Black or African American in Segment

The categorical percent black or African American in segment 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 2000 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

Table 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
1 (12-14, 15-17)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
2 (12-14, 18-25)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Education (older); Employment (older)	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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Education (older); Employment (older)

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

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
3 (15-17, 15-17)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Number in Household Aged 0-11; Number in Household Aged 12-17; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
4 (15-17, 18-25)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Education (older); Employment (older)	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); Gender (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Education (older); Employment (older)
5 (18-20, 18-25)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Marital Status (younger); Education (older); Education (younger); Employment (older); Employment (younger)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Marital Status (older); Marital Status (younger); Education (older); Education (younger); Employment (older); Employment (younger)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Education (older); Employment (older); Employment (younger)

Table 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Marital Status (older); Marital Status (younger); Education (older); Education (younger); Employment (older); Employment (younger)	Race (older); Gender (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Number in Household Aged 12-17; Race (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
7 (12-14, 26+)	Race (older); Categorical Percent Hispanic or Latino in Segment; Age Category (older); Education (older)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)
8 (15-17, 26+	Race (older); Gender (older); Gender (younger); Census Region; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)

Table 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+)	Race (older); Gender (older); Gender (younger); MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (younger)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Marital Status (younger); Education (older); Employment (older); Employment (younger)	Number in Household Aged 35-49; Number in Household Aged 50+; Gender (younger); MSA; Categorical Percent Black or African American in Segment; Age Category (older); Marital Status (older); Marital Status (younger); Employment (older); Employment (younger)
10 (21+, 26+)	Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Age Category (older); Education (older); Education (younger)	Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Education (younger); Employment (older); Employment (younger)	Number in Household Aged 35-49; Number in Household Aged 50+; Race (older); Gender (older); Gender (younger); Census Region; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Age Category (older); Education (older); Employment (older); Employment (younger)

Q.3 Multiplicities

Table Q.2 Model Summaries (Multiplicities)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
Parent-Child (12-20) Child Focus	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)	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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status (older); Education (older); Employment (older)
Parent-Child (12-20) Parent Focus	No model used; no nonrespondents	No model used; no nonrespondents	No model used; no nonrespondents
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment

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

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
Sibling (12-14) Sibling (15-17) Younger Sibling Focus	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-17) Sibling (18-25) Older Sibling Focus	MSA; Categorical Percent Black or African American in Segment	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-17) Sibling (18-25) Younger Sibling Focus	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status; Education (older); Employment (older)	Household Size; Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status; Education (older); Employment (older)	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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment; Age Category (older); Marital Status; Education (older); Employment (older)

Q.4 Household-Level Person Counts

Table 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 Younger Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Parent-Child (12-20) Child Focus, At Least One Pair Member Older Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Parent-Child (12-20) Parent Focus, Both Pair Members Younger Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment

Table 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	Variables Included in Predictive Mean Model: Not Including Household Size
Parent-Child (12-20) Parent Focus, At Least One Pair Member Older Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Race (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment	Number in Household Aged 0-11; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Both Pair Members Younger Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, At Least One Pair Member Older Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment

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

Pair Domain	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 Younger Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, At Least One Pair Member Older Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Spouse-Spouse, Both Pair Members Younger Than 18	Gender (older); Gender (younger); Categorical Percent Hispanic or Latino in Segment; Categorical Percent Owner-Occupied in Segment	Age Category (older); Race (older); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American 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); Gender (older); Gender (younger); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment

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

Pair Domain	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; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Race (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment	Race (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Spouse-Spouse With Children, Both Pair Members Younger Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Spouse-Spouse With Children, At Least One Pair Member Older Than 18	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category (older); Race (older); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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); Gender (older); Gender (younger); Marital Status (older); Education (older); Employment (older); Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment

Table Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model: Including Household Size	Variables Included in Predictive Mean Model: Not Including Household Size
Parent-Child (12-20) Child Focus, Younger Than 18	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Parent-Child (12-20) Child Focus, Older Than 18	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Parent-Child (12-20) Parent Focus, Younger Than 18	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Parent-Child (12-20) Parent Focus, Older Than 18	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Number in Household Aged 0-11; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment

Table Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (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), Older Sibling Focus, Younger Than 18	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Older Than 18	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Younger Than 18	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment

Table Q.4 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Pair Domain	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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment
Spouse-Spouse, Younger Than 18	Household Size; Age Category; Race; Gender; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Age Category; Race; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Number in Household Aged 0-11; Age Category; Race; MSA;
Spouse-Spouse, Older Than 18	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Age Category; Race; Gender; Marital Status; Education; Employment; Census Region	Race; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment
Spouse-Spouse with Children	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied in Segment	Household Size; Age Category; Race; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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; Gender; Marital Status; Education; Employment; Census Region; MSA; Categorical Percent Hispanic or Latino in Segment; Categorical Percent Black or African American in Segment; Categorical Percent Owner-Occupied 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 was considered the direct count, and the count from the other pair member was considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided both rosters did not have bad data codes. The strategy also was usually clear if one pair member had bad data in the household roster, or had a 0 count when the pair relationship precluded a value of 0. The count from the pair member with good, nonzero data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes 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.

Note that the reconciliation of differing counts was necessary for parent-child and sibling-sibling pairs but was not necessary for spouse-spouse pairs, since the multiplicity count for spouse-spouse pairs was always 1. As noted in Section 6.3, it was technically possible for a respondent to have multiple spouses, but these situations were not accounted for.

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 in most cases had values of 1 or 2. If neither side had bad relationship codes and the direct count exceeded the indirect count, 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. This was done in two ways. First, the total number of roster members between ages 30 and 39, 35 and 44, and 40 and 49 were compared between pair members and the screener. The pair member with age range counts closest to the screener was the one whose parent-child count was chosen for the final count. If neither side had age range counts equal to the screener, then the pair member with a parent-child count equal to the total number of screener roster members between ages 26 and 64 was chosen as the final count.

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 or 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 exceeded by the indirect count, then the child listed only one parent, and the parent listed a spouse (a "stepparent") or live-in partner in the household roster. The following rules applied:

1. The indirect count might have exceeded the direct count because the selected child did not accept a stepparent or live-in partner as his or her parent. If this stepparent or live-in partner was the other respondent selected, we determined that 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 was not 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 or 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 identified only 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 did not have since the live-in partner was not selected. Hence, these cases were left to imputation.

3. If age range counts between the two pair members and the screener matched across a variety of age ranges (30 to 39, 40 to 49, and 50 to 59), but the child's roster had a bad relationship code among roster members of potential parent age (15 or older), or the child's roster had a value of MBRSEL that did not match what was finally determined to be the child's parent, then the multiplicity count for the parent—the indirect count—was selected as the final count.

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. 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, then the multiplicity count that corresponded to the pair member with the same count of children as the screener was used.

2. 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, then 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.

3. The direct count and indirect count might differ because either the child listed a sibling that the parent considered "another relative" or the parent listed a child that the child considered "another relative." In either case, the parent was the one to answer the FIPE3 question. Because of this, the multiplicity count from the parent's perspective was selected as the final count, provided that the counts of children in the household within the relevant age ranges for each pair member were equal.

4. After considering the above situations, the multiplicity counts might have still differed without a resolution of which count should have been chosen as the final count. This occurred because the counts of children in the household differed between pair members, each of which differed from the screener count. Moreover, multiplicity counts did not exceed the screener age range count. In this instance, if one of the multiplicity counts equalled the screener age range count, then this multiplicity was selected as the final count.

Because of the hierarchical nature of these counts, parent-child counts for 12- to 17-year-old and 12- to 20-year-old children could sometimes be derived if the 12- to 14-year-old parent-child count was already determined for both child focus and parent focus counts. In particular, if one pair member's count for 12- to 17-year-old children or 12- to 20-year-old children equalled or exceeded the final parent-child count for 12- to 14-year-old children and the other did not, then the pair member's count that equalled or exceeded the 12- to 14-year-old count was chosen 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). Conversely, the multiplicity count was the number of siblings in the older age group (15 to 17 or 18 to 25) when the younger sibling was the focus. Deciding how to assign a final multiplicity count often involved looking at a count of household members within the age range of the siblings being counted. For example, if the older sibling was the focus and the age ranges were 12 to 14 and 15 to 17, the number of household members aged 12 to 14 were counted. The following general rules applied if the multiplicity counts for each pair member disagreed:

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. In particular, depending upon the domain, the count of household members within the age range of the siblings being counted was compared between each pair member and the screener. The multiplicity count from the pair member with the count closest to the screener was used, provided that the member had no bad relationship codes within the relevant age range.

2. If the counts of household members within the age range of the siblings being counted differed between pair members and those counts were both exceeded by the screener count, then the multiplicity associated with the pair member with the age range count closest to the screener was chosen, provided that the member had no bad relationship codes within the relevant age range.

3. In some cases, the counts of household members within the age range of the siblings being counted were the same for the two pair members, but the multiplicity counts disagreed.

   a. If one pair member had bad relationship codes and the other did not, the disagreement could have been due to the bad relationship codes. If the sum of the multiplicity count and the number of bad relationship codes were equal across pair members, then the final count was set equal to the multiplicity of the pair member who did not have bad relationship codes.

   b. If one pair member identified the other as "sibling" but the other pair member did not reciprocate, then imputation was required to establish whether the relationship was sibling-sibling. The count associated with the pair member who indicated that the other pair member was a sibling should have been chosen as the final count. In effect, this was done by taking the maximum of the two pair members' counts.

4. If the counts of household members within the age range of the siblings being counted disagreed and both exceeded the screener count of household members within the relevant age range, then 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.

5. If differing multiplicity counts could not be reconciled with the above rules, upper and lower bounds for the true multiplicity were determined using the two multiplicity counts, as well as the counts of children within relevant age ranges in both pair member's rosters and the screener roster. In rare cases, the values for these bounds were equal. These cases were investigated, and if the reasons were legitimate, then the final multiplicity count was set to this value. Otherwise, the final multiplicity was left to imputation.

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 did not involve a reconciliation of counts, though assigning a final count meant ensuring that pair relationships were not hidden due to the relationships of the two pair members to other household members.[34] A similar situation occurred 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, provided pair relationships of interest were not hidden. If bad data existed in either household roster, but the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes 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 assign a final count, as well as to reconcile 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 older, 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 nonparent-child pairs of interest, in most cases, the counts of children in the relevant age range with parent(s) in the household (abbreviated below as children

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

with parent(s) in the household) 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:

- Either no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or the counts of children with parent(s) in the household 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, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household; and[35]
- The household size was greater than 1 and was nonmissing on both sides.

2. The counts of children with parent(s) in the household might have agreed even though the above conditions were not met. The final count of children with parent(s) in the household could still have been set to one of the sides, if any one of the following was true:

   - If the number of children within the relevant age ranges matched across both rosters and the screener and (at least) one side had all good age and relationship codes, provided the equal counts did not refer to the same children;[36]
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and (at least) one side had all good age and relationship codes;
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and the number of respondents who were old enough to be parents in the household was zero according to the screener; or
   - If the counts of children with parent(s) in the household that agreed with each other equalled or exceeded the count of the number of children from the screener within the relevant age ranges.

3. The counts of children with parent(s) in the household 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

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

[36]This was determined by excluding situations where the ages of the identified parents did not match, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household

included in each other's rosters, but they were obviously in the screener roster, so the final count of children with parent(s) in the household was set to 2.

4. If one pair member did not have a valid roster but the other member did, the final count of children with parent(s) in the household was set to the other pair member's count under the following conditions:

   - No counts of one child with parent(s) in the household when both members of the pair were children in the relevant age range and the children were siblings, and
   - Either:
     - There were no bad relationship codes within the relevant child-age ranges and the respondent identified parents in the household,
     - 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.

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 of children with parent(s) in the household, if that count was 0, 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,
   - The other roster also was 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 "grandchild" relationship code(s) were incorrectly entered into the respondent's household roster. The "grandchildren" that these relationship codes were referring to were not the respondent's grandchildren, but, rather, they were the respondent's grandparent's grandchildren.[37]

6. When two different family units were in the household, the determination of the final count of children with parent(s) in the household had to be treated separately. This could have included the multigenerational families referred to earlier and the 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 of children with parent(s) in the household (one count might be 0) was used as the final count, provided the following conditions were satisfied on both sides:

   - There were no bad ages or relationship codes within the relevant age ranges;

[37] This condition has not manifested itself since the 2001 survey. With the addition of a new consistency check added since the 2001 survey to address grandparent/grandchild code inconsistencies, this condition could be observed only if a respondent overrode this consistency check, which has not happened.

- Both had counts of children with parent(s) in the household pointing to two or fewer parents, meaning that the two family units were not identifiable on a side;
- The number of identified parents was not equal to the total number of household members older than 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 was 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; and
- 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 pair was a parent-child pair, the final count was determined using imputation.

7. Two family units might be in the household but the conditions given in item #6 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of children with parent(s) in the household were determined for each side. The maximum of the two counts was used as the final count if the household members in the roster older than 25 (of parental age) were either both equal to the number of household members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of children with parent(s) in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.

8. 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 of children with parent(s) in the household from his roster was only 1, then the final count was set to 2.

9. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, it was necessary to decide who to believe. This occurred often 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 the respondent was a

boarder.[38] Selecting either the zero count or nonzero count in this instance required that the following conditions were met:

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

When one count of children with parent(s) in the household was zero and the other was nonzero, the nonzero count was used under the following conditions:

- The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges, and
- 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, and the count of children with parent(s) in the household did not exceed the screener count of children within the relevant age range;
  - The count of children in the household within the relevant age range for the nonzero count pair member matched that of the screener;
  - 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;

    The respondent with a zero count had no household members with a family-type relationship code where the reported household sizes of the two pair members were equal (indicating that it was unlikely that anyone had entered or left the household between interviews);
  - The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within the relevant age range in the household for the zero count was closer to the screener age count. Nevertheless, the nonzero count was equal to or less than the screener age count; or

[38] 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.

- 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, 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 of children with parent(s) in the household was used if the zero-count respondent had no bad relationship codes at all, and either:

- 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.

10. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, taking the side that was closest to the screener sometimes meant that the count of children with parent(s) from neither pair member was chosen. As with the previous item, a zero count and a nonzero count often 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 the respondent was a boarder. If neither the zero count nor the nonzero count was chosen, the final count could still have been determined using either the screener count, the count of children within the relevant age range for the respondent with a zero count, or one less than the nonzero count. One of these was chosen, provided that the following conditions were met:

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

The screener count was chosen if either:

- The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges. 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, and the nonzero count exceeded the screener count of children within the relevant age range, or
- The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within

the relevant age range in the household for the respondent with the zero count was closer to the screener age count, and the nonzero count exceeded the screener count of children.

In situations where a respondent with zero count had a roster more closely resembling that of the screener, but the screener included a household member within the relevant age range who was not part of the immediate family, neither the nonzero count of children with parent(s) in the household nor the screener count of children within the relevant age range could be used—a different count had to be used. Two strategies were employed:

- For the respondent with a nonzero count of children with parent(s) in the household, the nonzero count was the same as the count of children within the relevant age range in the household, but it exceeded the number-of-children count for the zero-count respondent. However, the count of children within the relevant age range for the zero-count respondent, which was not zero, was closer to the screener age count than the nonzero-count respondent.
- If the count of children within the relevant age range for the zero-count respondent was the same as the nonzero count of children with parent(s) in the household, the number-of-children count for the zero-count respondent could not be used, since the nonzero count included a household member that was not in the appropriate age range at the time of screening. One less than the nonzero count of children with parent(s) in the household was therefore chosen as the final count.

11. Other situations with a zero and nonzero count did not necessarily mean that the relationship was something other than parent-child or sibling-sibling. 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 relevant child age range for the pair member with bad relationship codes in the roster, the nonzero count was selected.

12. The two counts of children with parent(s) in the household might have disagreed where both were nonzero and both exceeded the screener count of children within the relevant age range. For the screener count to be chosen as the final household count of children with parent(s) in the household, the following conditions had to be met:

    - The pair member's household rosters had to have different numbers of children within the relevant age range,
    - The pair relationship could be neither parent-child nor sibling-sibling with a zero screener count of children within the relevant age range(s),
    - The total number within the screener roster (where the minimum age was 12 years) had to be at least two, and

- The number of children in the screener roster within the relevant age range was valid and at least as large as the final count of children with parents in the household for the next smallest age range.

13. 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 equalled the larger count, the larger count was chosen.

14. The two counts might have disagreed because they disagreed on the ages of one or more household members, 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.

15. The two counts might have disagreed because they disagreed on the ages of one or more household members 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 1 but less than the other, then the screener count was used as the final count.

16. 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 were obtained using the parent-focus multiplicity counts corresponding to the appropriate age range. 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, then the nonzero count was used as the final count.

17. 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.

18. 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 0, it necessarily means that the parent-focus counts will also be 0. 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 nonparent-child pairs of interest, in most cases, the counts of parents with children in the household 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 a parent,
- The household size was greater than 1 and nonmissing on both sides, and
- Either:
    - No bad relationship codes for household members of an age to be parents,
    - The total count was 2 for two parents, or
    - The total count plus the number of grandparents equalled the total number of household members aged 26 or older, according to the screener roster.

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 of parents with children in the household might 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 of parents with children in the household 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 0, which would necessarily be the case if the child-focus counts were 0.

4. When two different family units were in the household, the determination of the final count of parents with children in the household had to be treated separately. This could have included multigenerational 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 count might be 0) was used under the following conditions:

    - There were no bad ages or relationship codes within the relevant age ranges,
    - Both pair members 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 was not equal to the total number of household members older than 25 on either side, meaning that parents could correspond to roster members identified by other relationship codes, and
    - 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.

5. Two family units might be in the household but the conditions given in item #4 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of parents with children in the household were determined for each side. The maximum of the two counts was used as the final count if the household members older than 25 (of parental age) in the roster were either both equal to the number of members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of parents with children in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.

6. 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 older.[39]

7. The two counts might have disagreed with one nonzero count and the other equal to zero. Due to the fact that the counts of parent(s) in the household with children were determined first and that the zero counts were handled separately, the final count of parents with children in the household determined at this stage of processing had to be nonzero. Counts arising from two or more families in the household also were handled in previous code. Hence, the final count had to be one or two parents.[40] The nonzero count was chosen as the final count if one of the following conditions were met:

    - 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.

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

    - The difference between the screener count of the number of household members aged 26 or older and the pair members' counts of this number of household members was not the same between the two pair members,

[39] For this condition, either the count for the other pair member was 0 or the count for the pair members was equal.

[40] This precluded the extremely unlikely possibility that the pair member with a zero count masked a situation where three parents in a single family unit lived in the household (two biological parents and a stepparent).

- Neither pair member had bad ages in their rosters, and
- 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 aged 26 or older.

9. The two counts might have 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.

10. 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 two, so the final count was set to 2.

11. 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 were obtained using the child-focus multiplicity counts corresponding to the appropriate age range.

12. 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.

13. 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 lower age range was 12 to 14 or 12 to 17 or whether the upper 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 of interest are of roster members in the upper 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 upper age range, and could not have a younger sibling in the lower 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 have been set to one of the sides if any one of the following conditions was true:

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

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 0.
    - Either:
        - There were no bad relationship codes within the upper age range,
        - The count was equal to the screener age count within the upper age range, or
        - The count was 0, and the count of household members in the lower age range was 0.

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 0,
    - The number of children in either the lower or upper age ranges was 0 with no bad ages in the roster

5. If neither pair member had a valid roster, it was occasionally still possible to assign a final count. If the number of children in the screener roster in either the lower or upper age ranges was zero and the screener roster was valid, then it was not possible for a sibling-sibling pair in the relevant age ranges to be selected and the final count to be set to 0.
6. 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 count might be 0) was used, provided the following conditions were satisfied for both pair members:

- The sum of counts of the number of sibling-sibling pairs equalled or exceeded at least one of the counts of household members in the upper age range for the screener roster or either of the pair member's rosters.
- 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.

7. 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, then the side with the number in the lower age range equal to the number in the screener roster within the lower age range was chosen. (In all cases, one side had a 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 roster 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 final 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 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 lower 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 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.

8. 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, then the count for the pair member with the good roster was selected.

9. If the counts from the two pair members did not agree, and 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 0. 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;
     - The other pair member had at least one relationship code indicating a relationship other than parent, child, sibling, spouse, grandchild, or grandparent;
     - 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.

10. If one pair member had no bad relationship codes within both the upper and lower age ranges, but the other member 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.

11. 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), and 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 a nonzero count was selected. If the number of household members within that age range in the screener roster was nonzero, then that number was chosen as the final count. Otherwise, the number of household members within the upper age range for the pair member with nonzero count was selected as the final count.

12. 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) in the roster for the pair member with the zero count was less than his or her total household size.

13. 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 a 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 a 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.

14. If the pair relationship was sibling-sibling and the sibling-sibling counts were associated with the same age range, then the household-level person counts were obtained using the younger sibling-focus multiplicity counts corresponding to the appropriate age range.

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.[41] If the household size was one, or the number of respondents aged 15 or older in the household was one or zero, then the final household person count was set to 0 since no spouse-spouse pairs could reside under those limits. If two family units had been previously identified in the household, the following rules were used to determine the final household person count:

1. 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 count might be 0) 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 children younger than 18 years of age. If two family units were multigenerational families, then the final count was obtained by taking the maximum of the two pair members' counts.

2. 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:

[41] In rare cases, an individual might identify two spouses in the household. As noted in Section 6.3, the true multiplicity count in these cases was not determined; rather, the multiplicity count was set to 1, due to the complexity of determining the appropriate multiplicity count and the rarity of the occurrence of multiple spouses.

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

3. If the conditions for the previous item were not met, it was still possible for two different spouse-spouse pairs to be in the household, even though two different family units were not previously identified. The final count was set to 2 under the following conditions:

   - One pair member had two parents with valid ages and both ages differed from the age of the spouse of the other pair member, and
   - The pair relationship was either sibling-sibling or a pair that was not a pair of interest.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children, grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition—the final count was assigned using the following rules:

4. 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 aged 15 or older for either pair member,
   - The number of spouse-spouse pairs was either one or zero for both pair members,
   - The household size was greater than 1 and nonmissing on both sides,
   - One pair member had at least two household members aged 15 or older, and
   - There were not two spouse-spouse pairs in the household according to the conditions given in item #3.

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

- One pair member was younger than 18 and had no bad relationship codes for roster members aged 18 or older, but he or she did have bad relationship codes for roster members between the ages of 15 and 17 years old. The other pair member had no bad relationship codes for roster members aged 15 or older.
- One pair member had a single bad relationship code, and no other relationship codes could match it to make it a couple (i.e., the pair member did not have a single identified parent, grandparent, parent-in-law, or child-in-law). The other pair member had no bad relationship codes.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where the pair member with no bad roster entries had the same number of household members aged 15 or older as the screener. The pair member with the bad roster entries would have had the same age composition as the screener if the number of roster members aged 15 or older was added to the number of roster members with bad ages.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where all the relationship codes for the pair member with no bad roster entries were immediate family codes (child, parent, sibling, spouse, partner, grandparent, grandchild). For the pair member with bad roster entries, all the existing relationship codes were immediate family codes

6. 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 1 if any one of the following conditions was true:

   - Both sides had fewer than four people older than 15 in the household, or
   - One side had fewer than four people older than 15 in the household, and the other side had no bad relationship codes among roster members aged 15 or older

7. 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 any one of the following conditions:

   - There were no bad relationship codes among roster members aged 15 or older, or
   - There were no bad relationship codes among roster members aged 18 or older and the pair member had parents.

8. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could have been 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:

- The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the smaller spouse-spouse count.
- The screener count of roster members aged 12 or older was smaller than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.
- The difference between the screener count of roster members aged 12 or older and the count of roster members aged 12 or older 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 have been 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:

    - The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.
    - The screener count of roster members aged 12 or older was larger than the count of roster members aged 12 or older 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 a 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 did not identify a spouse, partner, two parents, or two grandparents. The following conditions were required to select the nonzero count:

    - The pair member with a nonzero count either identified a spouse, a partner, two parents, or two grandparents.
    - The number of roster members aged 15 or older 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 two parents, the following additional conditions were required:
        - The number of roster members between the ages of 26 and 44 was the same between the two pair members.
        - The number of roster members between the ages of 30 and 49 was the same between the two pair members.
        - The number of roster members between the ages of 35 and 54 was the same between the two pair members.

- The number of roster members between the ages of 40 and 59 was the same between the two pair members.

- If the side associated with the nonzero count identified two grandparents, the following additional condition was required:

  - The number of roster members aged 50 or older was the same between the two pair members.

11. The counts might not agree because a pair member's partner did not consider the other pair member’s family as his or her own family. If at least one side identified a partner and the maximum count was 1, then the maximum was selected if both pair members had the same number of household members aged 15 or older. Otherwise, if the pair members had a different number of household members aged 15 or older, the count belonging to the pair member with a count of household members aged 15 or older closer to that of the screener was used as the final count.

12. The counts might not agree because a pair member had two grandparents and an uncle/aunt husband-wife pair in the household. The maximum was selected if the pair member associated with the smaller count had a grandparent and had at least two roster members who were neither parents, siblings, children, spouses, partners, or grandparents, and the pair member with the larger count had children-in-law.

13. 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 if the pair member with the smaller count did not have a spouse or partner but did have siblings aged 15 or older, and there were household members in his or her roster that were not parents, children, siblings, spouses, partners, grandchildren, or grandparents.

14. 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 the following conditions were met:

    - The pair member's roster associated with the minimum count (usually 0) had no children, parents, siblings, spouses, partners, grandchildren, or grandparents among respondents aged 12 or older, and
    - The pair member's roster associated with the maximum count had some roster members who were not children, parents, siblings, spouses, partners, grandchildren, or grandparents.

    Note that 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.

15. The count of the number of spouse-spouse pairs might not agree because the pair were siblings, and one sibling did not consider a stepparent or parent's partner as a "parent." The maximum count was selected if the following conditions were met:

    - 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, grandparents) than the roster associated with the other pair member.

16. The count of the number of spouse-spouse pairs might not agree because the household 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:

    - 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.

17. 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 number of screener roster members aged 12 or older exceeded the corresponding count from the questionnaire rosters of both pair members, which also differed from each other.

18. The counts might not agree because, 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:

    - The pair was a spouse-spouse pair.

- The pair member with the maximum count had two 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 were in fact a spouse-spouse pair. There was no way to check whether a grandfather was the father's father and the grandmother was the mother's mother, for example.

19. 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:

    - The number of screener roster members aged 12 or older 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 aged 15 or older on either pair member's roster.

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

21. After accounting for all other rules, if the number of spouse-spouse pairs was still missing, but the lower and upper bounds for imputation were equal to each other, then the final household-level person count was set to one of those bounds.

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 first two rules described in this section were determined directly from the spouse-spouse counts or from the household size, and no reconciliation of counts was necessary:

1. For a sizable proportion of cases, clearly no couples with children could be in the household, either because the spouse-spouse count was 0 or the household size was two or less. In these cases, the final spouse-spouse-with-children count was set to 0.

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

member had a spouse-spouse-with-children count that 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.

The remainder of cases involved households with at least one spouse-spouse couple. After assigning values for the conditions described above, the assignment of values for these cases was done using the rules described in the rest of this section. If two family units had been previously identified in the household, the following rule was used to determine the final household person count:

3. 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 count might be 0) was used, provided the spouse-spouse count was greater than 1. In that event, the maximum count was used.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children, grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition (unless specifically noted below)—the final count was assigned using the following rules:

4. 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:

   - 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 younger than 18, and one of the following conditions held for at least one pair member:
       - The pair member's roster had no bad relationship codes for roster members aged 15 or older,
       - The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or
       - The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 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).

5. The pair members might both have had zero counts, but the above conditions did not apply. The final count could still have been 0 if the age counts for both pair members and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)

6. The counts for both pair members might still have agreed with nonzero counts, even though none of the previous conditions applied. The final count could still have been set to one of the pair member's counts if the pair relationship was imputed to be a spouse-spouse pair with children.

7. 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 one of the following conditions:

    - 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 younger than 18, and one of the following conditions held for the pair member with the valid roster. Either:
        - The pair member's roster had no bad relationship codes for roster members aged 15 or older,
        - The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or
        - The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 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).

8. The pair member with the valid roster might have had a zero count, but the above conditions did not apply. The final count could still have been 0 if the age counts for both the pair member with the valid roster and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)

9. If the spouse-spouse-with-children counts disagreed in the same manner as the spouse-spouse counts disagreed, then the choice was obvious: Use the count that corresponded to the correct spouse-spouse count. (In this case, it was not necessary to check for the potential of two or more family units in the household.) 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 count was set to one less than the final spouse-spouse count.

10. Based on earlier conditions, we already excluded households without couples. We 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 corresponded 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.
    - Either:
        - The relationship was parent-child where the child was between the ages of 12 and 17, or
        - The relationship was parent-child where the child was between the ages of 18 and 20 and the child had siblings younger than 18.

11. In some cases, two couples were identified in the household where the household was multigenerational (one member of the younger couple was 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 able to identify the nephews, nieces, brothers-in-law, or sisters-in-law—which 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 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 was in a marriage/partnership that occurred after an earlier marriage, the partner might not have considered the partner's children as his or her children, but the child (who also was selected) considered the spouse/partner a parent. Even though the pair relationship was not parent-child, these cases were still counted as spouse-spouse with children since they consisted of the children of one spouse/partner. The maximum count was selected under the following conditions:

- The pair relationship was not one of interest.
- One count was 0 and the other count was 1.
- The pair member with the zero count had a spouse or partner.
- The pair member with the nonzero count had parents.
- The spouse-spouse final count was nonmissing.

13. The counts might have been unequal because children younger than 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:

    - One pair member had a nonzero count of children younger than 18 and the other pair member had a zero count of children younger than 18.
    - Either:
        - The screener composition indicated that children younger than 18 were in the household, whereupon the nonzero count was selected, or
        - The screener composition indicated that no children younger than 18 were in the household, whereupon the zero count was selected.

14. The counts might have been unequal with a count of 0 and a count of 1 because a pair member with a count of 0 was not part of the immediate family unit. The nonzero count was used under the following conditions:

    - 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 aged 12 or older.

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

    - The pair member with a nonzero count was younger than 21 and had two parents.

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

- The number of roster members younger than 18 was the same between both pair members.
- The side with the smaller count had one bad relationship code for roster members younger than 18.

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.